

**Pontifícia Universidade Católica do Rio Grande do Sul
Faculdade de Informática
Programa de Pós-Graduação em Ciência da Computação**

**UM AMBIENTE INTEGRADOR
PARA ANÁLISE DE
PROCESSOS DE NEGÓCIO**

Cristian Tristão

**Dissertação apresentada como
requisito parcial à obtenção do grau
de mestre em Ciência da Computação**

Orientador: Prof. Dr. Duncan Dubugras Alcoba Ruiz

Porto Alegre
2006

Dados Internacionais de Catalogação na Publicação (CIP)

T838a Tristão, Cristian

Um Ambiente integrador para análise de processos de
negócio / Cristian Tristão. – Porto Alegre, 2006.

68 f.

Diss. (Mestrado) – Fac. de Informática, PUCRS

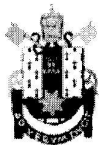
Orientador: Prof. Dr. Duncan Dubugras Alcoba Ruiz

1. Processos de Negócio – Gerência. 2. Processos de
Negócio - Análise. 3. Mineração de Dados. 4. Informática.

I. Título.


CDD 005.72

**Ficha Catalográfica elaborada pelo
Setor de Tratamento da Informação da BC-PUCRS**



TERMO DE APRESENTAÇÃO DE DISSERTAÇÃO DE MESTRADO


Dissertação intitulada "**Um Ambiente Integrador para Análise de Processos de Negócio**", apresentada por Cristian Tristão, como parte dos requisitos para obtenção do grau de Mestre em Ciência da Computação, Sistemas de Informação, aprovada em 14/12/2006 pela Comissão Examinadora:



Prof. Dr. Duncan Dubugras Alcoba Ruiz - PPGCC/PUCRS
Orientador



Prof. Dr. Osmar Norberto de Souza - PPGCC/PUCRS



Profa. Dra. Renata de Matos Galante - UFRGS

Homologada em 09/07/07, conforme Ata Nº. 16 pela Comissão Coordenadora.



Prof. Dr. Fernando Luís Dotti
Coordenador.

Aos meus pais, professores orientadores
e amigos... companheiros de todas as horas.

AGRADECIMENTOS

Aos meus pais pelo apoio e oportunidade.

A minha irmã pelo apoio, paciência e compreensão.

Ao Duncan, meu orientador, pela paciência, apoio, compreensão e amizade. Um verdadeiro paizão.

A professora Karin Becker, pela paciência, apoio, compreensão e amizade, um agradecimento especial. Sem palavras!

Aos colegas do curso de Mestrado em Ciência da Computação da PUCRS pela parceria, incentivo e amizade.

Aos professores e amigos que participaram de alguma forma deste trabalho.

Aos que não nomeei, mas que de alguma maneira participaram desta caminhada.

Enfim, a todos que colaboraram para que esse trabalho se realizasse, meu agradecimento sincero.

RESUMO

A análise de processos tem desempenhado um papel fundamental na gestão dos negócios. A maior parte das pesquisas e soluções existentes para essa análise focaliza muito exclusivamente na análise quantitativa de processos através de métricas de desempenho e indicadores de qualidade, possuindo menos recursos investigativos que permitam compreender o porquê do comportamento observado. Técnicas de mineração de dados possuem um papel importante neste contexto, oferecendo recursos para análise causal de comportamento. No entanto, as técnicas de análise e monitoração de processos de negócio são utilizadas de forma disjunta, ou seja, não existe uma integração ou complementação de informação entre as diferentes abordagens de descoberta de conhecimento. Além disso, não suportam a análise seqüencial dos fluxos de execução, prejudicando a investigação de anomalias e comportamentos específicos.

Este trabalho propõe um ambiente para a análise de processos de negócio com recursos, que combinam e permitem explorar, de forma sinérgica, as informações advindas da aplicação de técnicas de mineração seqüencial, originalmente propostas para a Mineração do Uso da *Web*, com aquelas sobre mensuração de processos, atividades e recursos.

Palavras-Chave: Gerência de Processos de Negócio. Análise de Processos de Negócio. Monitoração de Atividades de Negócio. Mineração de Dados.

ABSTRACT

The processes analysis has performed an essential role in businesses management. The majority of research efforts and existing solutions for this analysis focus very exclusively on the quantitative analysis of processes through performance metrics and quality indicators. Few of them investigate resources that allow understanding the reasons of the observed behavior. Data mining techniques have an important role in this context, offering resources for behavior causal analysis. However, the analysis and monitoring techniques of business processes have been a disjoint use. Indeed, there is not an information integration or synergy among different knowledge discovery approaches. Moreover they do not support the sequence analysis of execution flows, harming anomaly investigate and specific behaviors.

This work proposes an environment for the analysis of business processes taking into account related resources. Such environment permits the synergic exploration among information originated from the employment of sequential mining techniques, previously used by Web Usage Mining, with those from processes, activities and resources measurements.

Keywords: Business Process Management. Business Process Analysis. Business Activity Monitoring. Data Mining.

LISTA DE FIGURAS

Figura 1 – O ciclo de vida BPM comparado com WfMS (adaptação de [WES04]).	14
Figura 2 – Estrutura x métricas.	16
Figura 3 – Diagrama de atividades (pedido de pagamento).	19
Figura 4 – Proposta de arquitetura para BPA (adaptada de [GOL04]).	21
Figura 5 – Mineração e visualização de padrões de navegação.	24
Figura 6 – WebPath: interface de consulta (definição de critérios).	25
Figura 7 – WebPath: interface de visualização.	26
Figura 8 – Visualização de processos com Delta Analysis.	33
Figura 9 – Arquitetura do ambiente FlowSpy.	38
Figura 10 – Criação de perfis de análise a partir do modelo de um PN.	41
Figura 11 – Filtragem do log de execução.	43
Figura 12 – Abstração de visualização do tipo agregação.	44
Figura 13 – Abstração de visualização do tipo exclusão.	45
Figura 14 – Integração de informação.	47
Figura 15 – Diagrama de classes simplificado.	50
Figura 16 – Diagrama de casos de uso.	51
Figura 17 – Análise estrutural MUW → KPI.	53
Figura 18 – Análise quantitativa MUW → KPI.	54
Figura 19 – SQL para métrica da Figura 17.	54
Figura 20 – Análise quantitativa KPI → MUW.	56
Figura 21 – SQL para análise quantitativa.	56
Figura 22 – Análise estrutural KPI → MUW.	57
Figura 23 – Caso de uso definir perfil de análise.	58
Figura 24 – Caso de uso filtrar log de execução.	59
Figura 25 – Caso de uso abstrair visualização do tipo agregação.	61
Figura 26 – Caso de uso abstrair visualização do tipo exclusão.	62

LISTA DE QUADROS

Quadro 1 – Relação entre conceitos.	34
Quadro 2 – Quadro resumo dos trabalhos pesquisados.	36

LISTA DE SIGLAS

APM	Abstract Process Monitor
BAM	Business Activity Monitoring
BI	Business Intelligence
BPA	Business Process Analysis
BPC	Business Process Cockpit
BPI	Business Process Intelligence
BPM	Business Performance Management
CDW	Corporate Data Warehouse
CPMS	Corporate Performance Measurement System
CRM	Customer Relationship Management
DDS	Dynamic Data Store
DW	Data Warehouse
EAI	Enterprise Application Integration
EDW	Enterprise Data Warehouse
ERP	Enterprise Resource Planning
ETL	Extraction, Transformation e Loading
KDD	Knowledge Discovery in Databases
KPI	Key Performance Indicator
MUW	Mineração do Uso da Web
OLAP	On-Line Analytical Processing
OMG	Objected Management Group
PDW	Process Data Warehouse
PIF	Process Information Factory
PME	Process Mining Engine
PN	Processo de Negócio
RTI	Right-Time Integrator
SCM	Supply Chain Management
SQL	Structured Query Language
TI	Tecnologia da Informação
UML	Unified Modeling Language
WfMC	Workflow Management Coalition
WfMS	Workflow Management System

SUMÁRIO

1. Introdução	13
1.1. Caracterização do Problema	15
1.2. Objetivos	15
1.3. Estrutura do Trabalho.....	16
2. Fundamentação Teórica	18
2.1. Gestão de Processos de Negócio	18
2.2. Análise de Processos de Negócio.....	19
2.3. Mineração de Dados	22
2.3.1. Técnica Seqüencial MUW	23
2.4. Considerações	27
3. Trabalhos Relacionados.....	28
3.1. Business Process Intelligence.....	28
3.2. iBOM.....	30
3.3. Delta Analysis e Conformance Testing	31
3.4. Corporate Performance Measurement System	33
3.5. Process Information Factory	33
3.6. MUW x BPA	34
3.7. Considerações	35
4. FlowSpy	37
4.1. Arquitetura do Ambiente	37
4.1.1. Origem dos Dados	38
4.1.2. Modelo Analítico.....	39
4.1.3. Extrator de Dados	39
4.1.4. Perfil de Análise	40
4.1.5. Gestor de Abstração	42
4.1.5.1. Filtragem do Log	42
4.1.5.2. Abstração de Visualização	43
4.1.6. Mineração	45
4.1.7. Gestor de Métricas.....	46
4.1.8. Integrador.....	46
4.1.9. Visões	47
4.2. Considerações	47
5. Experimentação	49
5.1. Modelo de Dados Simplificado	50
5.2. Casos de Usos	51
5.2.1. Caso de Uso Mudar Análise MUW → KPI.....	52
5.2.2. Caso de Uso Alternar Análise KPI → MUW	55

5.2.3.	Caso de Uso Definir Perfil de Análise.....	57
5.2.4.	Caso de Uso Filtrar Log.....	58
5.2.5.	Caso de Uso Abstrair Visualização	60
5.3.	Considerações	62
6.	Conclusões e Trabalhos Futuros.....	63
	Referência Bibliográfica	65

1. INTRODUÇÃO

As empresas estão em constante busca por alternativas e soluções que possam reduzir os custos e maximizar os lucros. Para tanto, na década passada surgiu a idéia de gestão de processos de negócio ou BPM (acrônimo para *Business Process Management*). De acordo com a *Workflow Management Coalition*¹ (WfMC) [WMC99], Processos de Negócio (PN), ou simplesmente processo, é “um conjunto de um ou mais procedimentos ou atividades relacionados, os quais coletivamente atingem um objetivo de negócio, dentro do contexto de uma estrutura organizacional que define papéis funcionais e relações”. As primeiras abordagens de BPM objetivavam a automação e controle dos processos, proporcionando às empresas o conhecimento e domínio do funcionamento do seu negócio. Com o auxílio da Tecnologia da Informação (TI), diversas soluções e ferramentas foram desenvolvidas para gerenciar PN, e.g. *Oracle Workflow* [CHA02] e *Shark* [OPE06], que são ferramentas específicas de soluções da categoria *workflow*.

Workflow é uma das tecnologias mais utilizadas para a automação e controle de PN. *Workflow* é definido pela WfMC como a automação de processos de negócios, no todo ou em partes, na qual documentos, informações ou tarefas são passadas de um participante para outro, de acordo com um conjunto de regras procedurais. Um sistema de gerência de *workflow* (WfMS - *Workflow Management System*) é um sistema que define, cria e gerencia a execução de *workflows* através do uso de um software, executando um ou mais motores *workflow* [WMC99]. Um motor *workflow* é capaz de interpretar a definição do processo, interagir com participantes do *workflow* e, onde requerido, invocar o uso de ferramentas de Tecnologia da Informação (TI) e aplicações. Em resumo, um sistema de gestão de *workflow* é um tipo de sistema de gerência de PN.

Com o passar do tempo, o termo BPM tornou-se mais abrangente. As organizações não estão mais interessadas em apenas acompanhar e controlar o andamento de seus processos, mas sim em poder medi-los, associar valores à execução de cada instância de processo, atividade e recurso (tanto humano quanto computacional) e analisar seu desempenho [CAS05a] [GOL04]. Desta forma, a organização passou a poder verificar o comportamento do negócio e realizar modificações em seus processos, caso estas não atendessem aos objetivos estabelecidos. *Business Process Analysis* (BPA) [WES04], *Business Activity Monitoring* (BAM) [AAL03] [DRE03] [HEL02] [WES04], *Business Process*

¹ Entidade sem fins lucrativos, criada em 1993 com objetivo de aumentar a exploração da tecnologia *workflow*, através do desenvolvimento de padrões e terminologias comuns.

Intelligence (BPI) [GRI04] e *Business Performance Management* [GOL04], são denominações que representam esta nova abordagem de BPM. Por motivo de padronização, utilizaremos neste trabalho o termo BPA para representar a área de BPM destinada à análise e monitoração de PN.

Weske et al. [WES04] definem um ciclo de vida para BPM onde pode-se observar a diferença entre as abordagens de gestão e análise de PN. A Figura 1 (adaptada de [WES04]) ilustra este ciclo de vida. As fases de projeto, configuração e execução dos processos correspondem aos primeiros esforços de BPM (automação e controle), exemplificados por um sistema de gerência de *workflow* (WfMS). A fase de diagnóstico, que corresponde à análise dos dados de execução, representa a nova tendência de BPM (análise e monitoração).

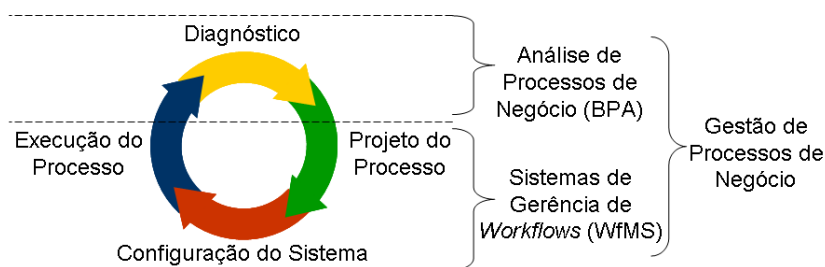


Figura 1 – O ciclo de vida BPM comparado com WfMS (adaptação de [WES04]).

Diante desta nova realidade, são apresentadas propostas para a gestão, análise e monitoração de PN, e.g. *Data Warehouse* (DW) específicos [CAS02] [LIS04] [SCH04], ambientes de BPM com suporte às novas necessidades [GOL04] [GRI04] e mineração de dados [AAL05] [CAS05b] [GRI04] para a descoberta de conhecimento. Propostas de BPA dizem respeito a sumarizações de dados históricos [CAS05b] [GOL04] [GRI04], monitoração das execuções através de indicadores de desempenho (KPIs) [CAS05b] [GOL04] [GRI04] e regras obtidas por mineração [AAL05] [AAL03] [GRI04]. No entanto, nenhuma destas abordagens utiliza mineração de dados para acompanhar de forma precisa a seqüência de atividades executadas por um conjunto de instâncias de processos, prejudicando a análise de ordem de execução de atividades e identificação de possíveis anomalias na execução do negócio ou problemas de modelagem do processo. Além disso, as técnicas de análise e monitoração de processos são utilizadas de forma disjunta, ou seja, não existe uma integração ou complementação de informação entre diferentes abordagens de descoberta de conhecimento.

Mineração do Uso da *Web* (MUW) é a área de pesquisa que visa extrair padrões de comportamento (navegação) do uso da *web* [COO99] [SRI00]. Spiliopoulou et al. [SPI00] [SPI98], seguindo esta linha, desenvolvem uma técnica de mineração de seqüências, a qual

possibilita a análise dos caminhos percorridos (seqüência de páginas acessadas). Com isto, é possível contrastar o uso esperado com o uso real de um *site*, analisando sua efetividade e *design*. Tristão et al. [TRI04] apresentam a ferramenta *WebPath* para a mineração e análise de padrões seqüenciais do uso da *web*, a qual implementa os algoritmos de mineração de seqüência propostos por Spiliopoulou et al. [SPI00] [SPI98]. Tristão et al. [TRI06] apresentam um estudo no qual utiliza-se *WebPath* no contexto de análise de PN. O objetivo desse estudo foi verificar a capacidade de obtenção de informações com o uso dessa técnica, sobre dados de execução de PN. Como resultado, obteve-se um conjunto de informações não apresentadas pelas demais técnicas (Capítulo 3) como, por exemplo, (a) o acompanhamento mais preciso da execução das atividades (b) a identificação de possíveis anomalias de execução e (c) o nível de convergência entre as atividades, i.e. qual a probabilidade de uma instância de processo seguir um ou outro caminho. Esse estudo também demonstrou uma limitação no que diz respeito à visualização de padrões envolvendo um grande número de atividades, sendo necessário complementá-la com mecanismos de abstração.

1.1. Caracterização do Problema

As primeiras propostas de BPA focam muito exclusivamente na análise quantitativa de processos através de métricas de desempenho e indicadores de qualidade. Atualmente estão sendo pesquisadas propostas possuindo recursos investigativos que permitam compreender o porquê do desempenho observado, e.g. técnicas de classificação (árvores de decisão) [GRI04] e de seqüência [AAL05]. No entanto, ainda existe uma carência de recursos para integrar informações de diferentes técnicas de mineração de dados e evidencia-se a falta de uma abordagem destinada à análise estrutural de PN para o acompanhamento detalhado dos fluxos de execução e que ofereça capacidades de análise mais específicas a visões do processo de negócio, ou que resultem em padrões de análise envolvendo apenas instâncias com características específicas às necessidades e interesse do usuário.

1.2. Objetivos

Este trabalho propõe um ambiente para a análise de processos de negócio que oferece recursos que combinem e permitam explorar, de forma sinérgica, as informações

advindas da aplicação de técnicas de mineração sequencial com aquelas sobre mensuração de processos, atividades e recursos. A idéia deste trabalho é ilustrada pela Figura 2. Dado fluxos de execução, verifica-se o seu desempenho mediante as métricas pré-estabelecidas que representem os objetivos da empresa. Da mesma forma, dado um conjunto de processos que apresentam um desempenho para suas métricas, tenta-se compreender o seu comportamento pela análise do fluxo de suas instâncias.

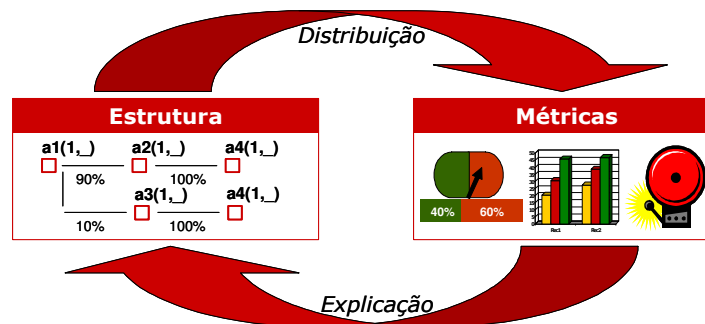


Figura 2 – Estrutura x métricas.

Visando obter informações relevantes com a aplicação de técnicas sequenciais de processos complexos, propõe-se a incorporação de mecanismos de abstração no ambiente sendo proposto, para prover ao usuário um suporte à análise de padrões de execução complexos, além de possibilitar diferentes visões do mesmo processo. Estes mecanismos objetivam auxiliar o analista na definição de atividades de interesse, podendo ser consideradas tanto para restringir as atividades que fazem parte do domínio de análise, quanto para visualizar os resultados, permitindo generalizar ou especializar as visões dos padrões de execução resultantes (função análoga aos mecanismos de OLAP).

São duas as contribuições desta pesquisa:

- Integração de informações oriundas de diferentes técnicas de KDD (*Knowledge Discovery in Databases*);
- Mecanismos que permitem aplicar a técnica de mineração sequencial a processos de negócio complexos, i.e. que envolvem um grande número de atividades e fluxos de execução.

1.3. Estrutura do Trabalho

Este documento está organizado da seguinte forma: o Capítulo 2 apresenta alguns conceitos, definições e informações relacionadas à gestão e análise de processos de negócio. O Capítulo 3 apresenta algumas propostas de trabalhos relacionados a este estudo mostrando

os avanços e limitações existentes. Uma proposta de ambiente de BPA, denominada *FlowSpy*, é apresentada no Capítulo 4 e nele são detalhados os serviços e funcionalidades de cada componente, dando destaque aos componentes propostos para possibilitar a sinergia de informação entre técnicas de análise (estrutural e quantitativa) e para facilitar a análise e interpretação dos fluxos de execução de instâncias de processos de negócio. O Capítulo 5 apresenta experimentações onde são detalhadas, na forma de casos de uso, as funcionalidades dos componentes propostos neste trabalho e apresentados no capítulo anterior. Por fim, o Capítulo 6 apresenta uma visão geral de todo o trabalho, expondo conclusões e sugestões para trabalhos futuros.

2. FUNDAMENTAÇÃO TEÓRICA

2.1. Gestão de Processos de Negócio

Gestão de processos de negócio, ou *Business Process Management* (BPM) diz respeito à automação, descrição, controle, análise e monitoração de PN. Aalst et al. em [AAL03] definem BPM como sendo “um suporte ao processo de negócio usando métodos, técnicas e software para projetar, representar, controlar e analisar processos operacionais envolvendo pessoas, organizações, aplicações, documentos e outros recursos de informação”.

Workflow é uma das tecnologias mais utilizadas para a automação e controle de PN. Para o estudo e compreensão deste assunto, deve-se levar em consideração alguns conceitos relacionados à tecnologia *workflow*. De acordo com a WfMC [WMC99]:

- Sub-processo: é um processo iniciado ou chamado por outro processo ou sub-processo, sendo responsável pela execução de parte do processo maior.
- Atividade: uma descrição de uma parte de trabalho que dá forma a uma etapa lógica dentro de um processo. Pode ser categorizada como manual ou automatizada.
- Instância (de processo/atividade): é a representação de uma execução de um processo, ou de uma atividade dentro de um processo, com os seus dados associados. Cada instância tem sua identidade visível externamente e um estado interno próprio e representa uma linha separada de execução de processo ou de atividade, podendo ser controlados de forma independente.
- Participante do *workflow* ou Ator ou Recurso: é quem executa o trabalho representado por uma instância de atividade de um *workflow* (é aquele que atua). Recursos podem ser pessoas ou processos automatizados.

Existem inúmeras formas de modelagem de *workflow*, sendo redes de Petri (*Petri Nets*) e diagrama de atividades UML² os mais utilizados. A Figura 3 ilustra o processo de negócio “pedido de pagamento” e sua representação, de acordo com o diagrama de atividades UML 2.0. Neste exemplo, inicialmente o banco deve receber um pedido de pagamento (início da atividade). Após o recebimento do pedido, o banco pode rejeitá-lo (quando o pedido é simplesmente fechado, e o fluxo encerrado) ou aceitá-lo. No caso de aceitação o pedido de

² *Unified Modeling Language* (UML) é uma especificação definida pela OMG destinada a modelar não apenas estrutura, comportamento e arquitetura de aplicações, mas também processos de negócio e estrutura de dados.

pagamento deve ser preenchido corretamente e após o término desta atividade, duas outras atividades são iniciadas: envio do pedido de pagamento e o envio da fatura. Estas duas atividades são executadas paralelamente de forma assíncrona. No primeiro fluxo, apenas deve ser enviado o pedido de pagamento. Já no segundo fluxo, a ação de envio de fatura gera o artefato fatura e a partir deste artefato é habilitada a realização do pagamento desta fatura. Após o pagamento, deve-se esperar a confirmação do pagamento. O pedido de pagamento é finalizado como pago, só após o término dos dois fluxos de execução.

Além deste(s) fluxo(s) básico, a qualquer momento entre a execução das atividades receber pedido, preencher pedido e enviar pedido, o processo de pedido de pagamento poderia ser cancelado e todo processo finalizado (representado pela “InterruptibleActivityRegion”) caso houvesse identificado alguma irregularidade ou recebido uma ordem explícita de cancelamento por parte de algum setor do banco competente a esta questão. Apenas ao nível de exemplificação, um dos fluxos de execução foi representado pelo conector ActivityEdge “A”.

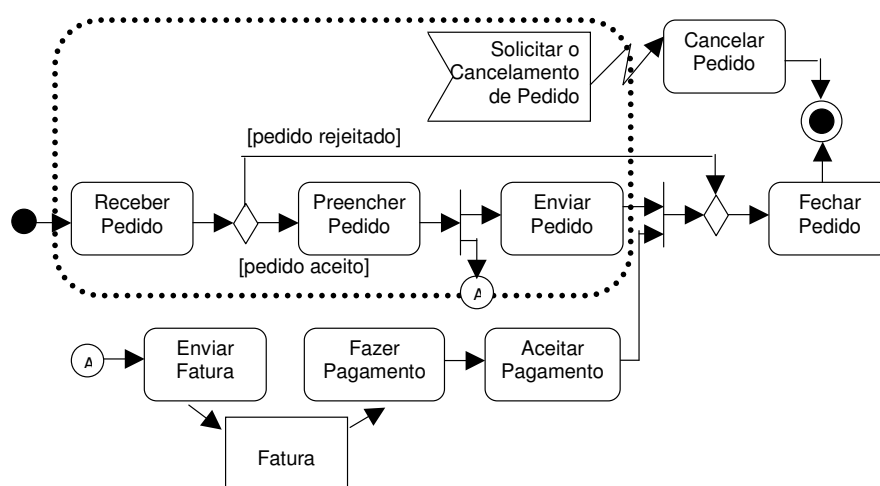


Figura 3 – Diagrama de atividades (pedido de pagamento).

(Fonte: [OMG04])

2.2. Análise de Processos de Negócio

Aos poucos, foi percebendo-se a necessidade de não apenas automatizar e controlar, mas também de medir, analisar e monitorar os PN. Atualmente, o principal objetivo é obter algum tipo de conhecimento útil sobre os dados gerados pelas ferramentas de controle de PN de modo a reduzir o tempo e o custo do produto ou serviço, e aumentar a qualidade e satisfação do cliente. Tendo em vista esta nova necessidade de mercado, vem surgindo um

novo campo de pesquisa conhecido como *Business Process Analysis* (BPA) [GRI04] [WES04]. Para Weske et al. [WES04], BPA é uma importante área de pesquisa que aborda a análise de processos de negócio, aspecto negligenciado pelas ferramentas de *workflow* tradicionais, indo de encontro às limitações das soluções tradicionais de BI (*Business Intelligence*).

Além de obter conhecimento útil e válido, Melchert et al. [MEL04a] [MEL04b] afirmam que o principal desafio de BPA está relacionado com a análise em tempo real, ou seja, reduzir a latência (período) do tempo entre a ocorrência de um evento do negócio que requeira uma ação apropriada pela organização e a tomada de decisão realizada. Segundo Melchert et al. [MEL04a] [MEL04b], a análise em tempo real é compreendida pela redução de três tipos de latência:

- *Latência de Dados* – tempo entre a ocorrência do evento do negócio e o armazenamento dos dados apropriados para a análise.
- *Latência de Análise* – tempo gasto para a realização da análise de dados e armazenamento de informação, desde a disponibilização dos dados.
- *Latência de Decisão* – tempo entre a entrega de informação e a tomada de decisão propriamente dita. Este tipo de latência depende muito do tipo de dados, da forma de visualizá-los e o tempo que os responsáveis têm disponível para analisar, decidir e executar suas decisões.

Golfarelli et al. [GOL04] afirmam que a redução destas latências não é considerada pelas soluções tradicionais de BI. Tendo em vista as necessidades e desafios encontrados na análise e monitoração de PN, Golfarelli et al. [GOL04] propuseram uma arquitetura completa para a análise e monitoração de PN. Nesta arquitetura, representada pela Figura 4, pode-se identificar os componentes responsáveis pela redução das latências de dados, análise e decisão.

Na redução da *latência de dados* as principais dificuldades dizem respeito ao controle de dados e interação entre atividades, à interpretação de eventos que acontecem em diferentes aplicações relacionados a um mesmo processo e à integração de dados. Isto porque muitos processos possuem uma infra-estrutura de suporte à informação composta por uma variedade de aplicações e de ambientes heterogêneos e distribuídos (e.g. CRM, SCM, ERP, EAI e *workflow*) [HEL02]. Técnicas de KDD estão sendo aplicadas para a integração de dados discrepantes e distribuídos, mais precisamente as etapas de conhecimento do domínio da aplicação, criação de um conjunto de dados (seleção), limpeza de dados e pré-processamento e transformação dos dados (ETL - *Extraction, Transformation e Loading*).

Alem destes, integradores de dados em tempo real (RTI – *Right-Time Integrator*) e armazenamento dinâmico de dados (DDS – *Dynamic Data Store*) são propostas para a redução da latência de dados. Trabalhos como [CAS05b] [CAS02] [LIS04] [SCH04] abordam a criação de DW e bases analíticas de dados para a unificação e disponibilização dos dados.

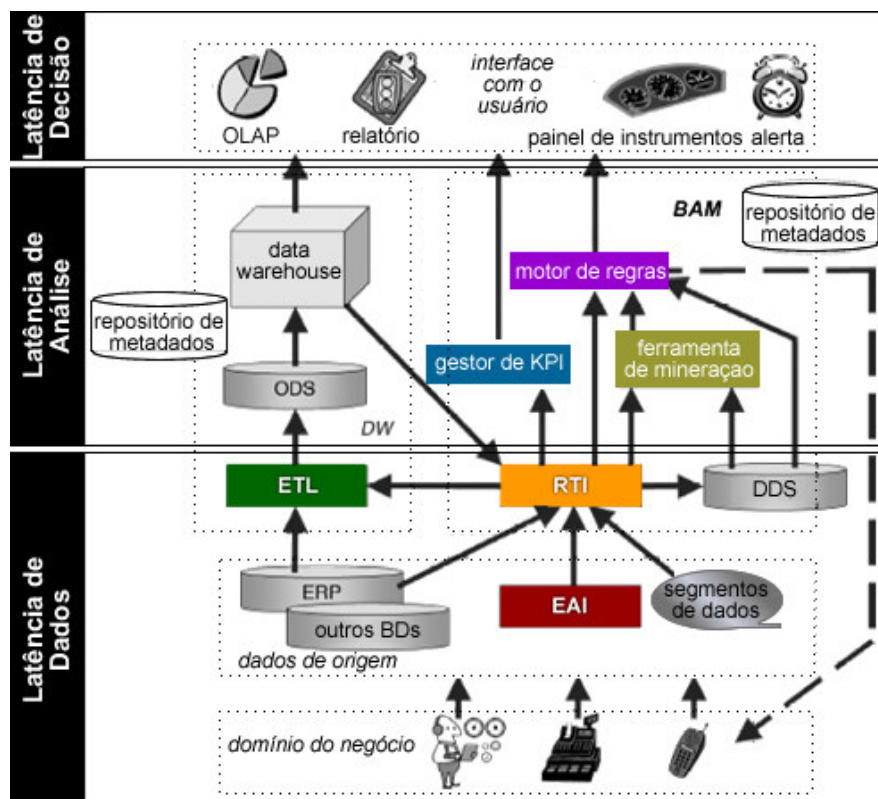


Figura 4 – Proposta de arquitetura para BPA (adaptada de [GOL04]).

Na redução da *latência de análise* o principal desafio está em proporcionar às empresas um melhor entendimento da situação de seus negócios e um aumento no poder de decisão sobre os processos. Para isto são utilizadas tanto as técnicas e ferramentas de BI tradicionais, tais como mineração de dados e ferramentas OLAP que atuam sobre dados históricos armazenados em bases de dados relacionais ou em DW modelados para análises específicas, quanto as novas tendências de BI, identificadas na Figura 4 por BAM, que possibilitam definir métricas (*KPI manager*) de acordo com os objetivos da empresa e medir continuamente os processos através de indicadores de desempenho (KPIs – *Key Performance Indicators*), ferramentas e técnicas de mineração sobre dados dinâmicos (*rule engine*). A Seção 2.3 apresenta algumas técnicas de mineração de dados. Destaca-se a técnica de mineração seqüencial para obter conhecimento sobre os fluxos de atividades.

O principal objetivo da redução da *latência de decisão* é dar acesso à informação gerada, de modo a suportar diferentes tipos de visualizações, filtrar e disponibilizar diferentes

visões sobre os mesmos dados do processo segundo diferentes óticas da organização. Isto porque, normalmente a empresa está dividida em diferentes níveis organizacionais, e cada nível está interessado em diferentes objetivos ou perspectivas de análise. Por exemplo, o diretor da empresa pode estar interessado em verificar o gasto mensal ocasionado por um conjunto de instâncias de um determinado processo e um gerente de projeto pode querer analisar o tempo gasto por cada recurso humano presente nestas execuções. Golfarelli et al. citam em [GOL04] que a estrutura de uma organização é normalmente dividida em 3 níveis hierárquicos: nível estratégico, onde é definida a estratégia global da empresa, e níveis tático e operacional, nos quais definem seus objetivos levando em consideração as metas estabelecidas na estratégia global (análise *top-down*). Por outro lado, informações e indicadores de desempenho coletados em níveis inferiores podem ajudar no ajuste e tomada de decisão pelos níveis superiores (análise *bottom-up*).

2.3. Mineração de Dados

A descoberta do conhecimento em base de dados ou KDD (*Knowledge Discovery in Databases*) é o processo não trivial de identificação de padrões que sejam válidos, novos, potencialmente úteis e compreensíveis [FAY96]. O processo de KDD refere-se ao conjunto de etapas que produzem conhecimentos a partir de dados e, entre elas, a mineração dos dados (*data mining*) que é a etapa que transforma dados em padrões [FAY96]. O processo KDD envolve vários passos, resumidos como [FAY96]: (a) conhecimento do domínio da aplicação, (b) criação de um conjunto de dados (seleção), (c) limpeza de dados e pré-processamento, (d) transformação dos dados, (e) mineração de dados, (f) análise e interpretação e (g) utilização do conhecimento obtido.

No processo de BPA todas estas etapas de KDD devem ser levadas em consideração para que se possa obter algum tipo de informação válida e útil pela transformação dos dados referentes à definição e execução dos processos de negócio.

Algumas técnicas de mineração de dados têm sido mais utilizadas do que outras para a análise de PN. A escolha de uma técnica de mineração depende muito do tipo de informação que se deseja obter. São alguns exemplos de técnicas de mineração usadas para a análise e monitoração de PN [HAN01]:

- Associação - objetiva encontrar conjuntos de itens que ocorrem juntos em uma base de dados e, a partir deles, formar regras de associação. As regras de

associação são representadas na forma $X \rightarrow Y$ (lê-se X implica Y). Em processos de negócios, regras de associação podem mostrar, por exemplo, que quando uma atividade X faz parte de um determinado processo, em 90% dos casos a atividade Y também aparece no mesmo processo. Este tipo de relação pode ser estabelecido entre recursos, tempo de execução de um processo e/ou atividade, etc.

- Classificação - objetiva classificar casos em diferentes classes, baseado em propriedades (atributos) comuns entre um conjunto de objetos em uma base de dados. O modelo de classificação construído é utilizado para prever classes de novos casos que serão incluídos em um banco de dados. Em processos de negócios, modelos de classificação podem ser construídos para caracterizar ou prever, por exemplo, se um processo será concluído ou não.
- Agrupamento - diferentemente da classificação, no agrupamento as classes de dados não são conhecidas, ou seja, a aprendizagem é não supervisionada. O processo de agrupamento visa a criação de classes. Em processos de negócios, o agrupamento, pode ser aplicado, por exemplo, para agrupar atividades, processos ou recursos, de acordo com características similares.
- Seqüência - o objetivo primário é encontrar todos os padrões seqüenciais com um suporte mínimo pré-definido, isto é, a porcentagem de seqüências de dados que contêm o padrão que está acima de um limite fornecido. Os padrões identificados são representados na forma de seqüências que refletem a ordem cronológica das ocorrências. Em análise de PN, a técnica de mineração seqüencial normalmente é utilizada para a descoberta do modelo de processos e alinhamento do negócio, conforme trabalho apresentado por Van der Aalst [AAL05]. Tristão et al. [TRI06] propõem o uso dessa técnica para a compreensão e análise detalhada de fluxos de execução (seqüência de atividades).

2.3.1. Técnica Seqüencial MUW

A Mineração do Uso da *Web* (MUW) [SRI00] visa extrair padrões de comportamento no uso de sites *web*. Ela tem como principal fonte de dados o *log* do servidor *web*, que registra todos os acessos às páginas feitos por usuários. Spiliopoulou et al. [SPI00] [SPI98] propuseram uma técnica de mineração de seqüências voltada à MUW, a qual possibilita a análise dos caminhos percorridos (seqüência de páginas) em sites. Assim, é

possível contrastar o uso esperado e o real de um site, analisando sua efetividade e projeto. Tristão et al. [TRI04] apresentam *WebPath*, uma ferramenta de mineração e visualização de padrões do uso da *web*, a qual disponibiliza a técnica de mineração de seqüências proposta por Spiliopoulou et al. [SPI00] [SPI98] para possibilitar a análise dos caminhos percorridos (seqüência de páginas) através de uma interface orientada a formulários para que o usuário especifique as propriedades dos padrões de navegação sobre os quais tem interesse.

Na técnica de mineração seqüencial do uso da *web*, os fluxos de execução (seqüência de páginas) são obtidos a partir de um *log* de execução previamente preparado, contendo todas as páginas acessadas pelos usuários, e posteriormente organizados em uma estrutura de árvore do tipo *trie*, denominada árvore agregada. Por exemplo, considere um *log* contendo 6 tipos diferentes de seqüências de páginas, doravante chamadas de fluxos, conforme esboçado na Figura 5.I. O número entre parênteses indica quantas sessões seguiram por um determinado fluxo. A árvore agregada (Figura 5.II) organiza esses fluxos unificando aqueles que possuem páginas em comum (e.g. iniciar pela página “a” ou pela página “b”). Ela possui um nodo raiz (^) que representa o número total de sessões analisadas e os demais nodos (páginas) são representados na árvore por uma tripla [P, O, A], onde P representa a página visitada, O a ocorrência desta página no fluxo e A o número de acessos desta ocorrência de página (e.g. o nodo [b; 2; 6] significa a segunda visita da página ‘b’ no fluxo, observado em 6 sessões). Arestas entre os nodos descrevem os caminhos percorridos.

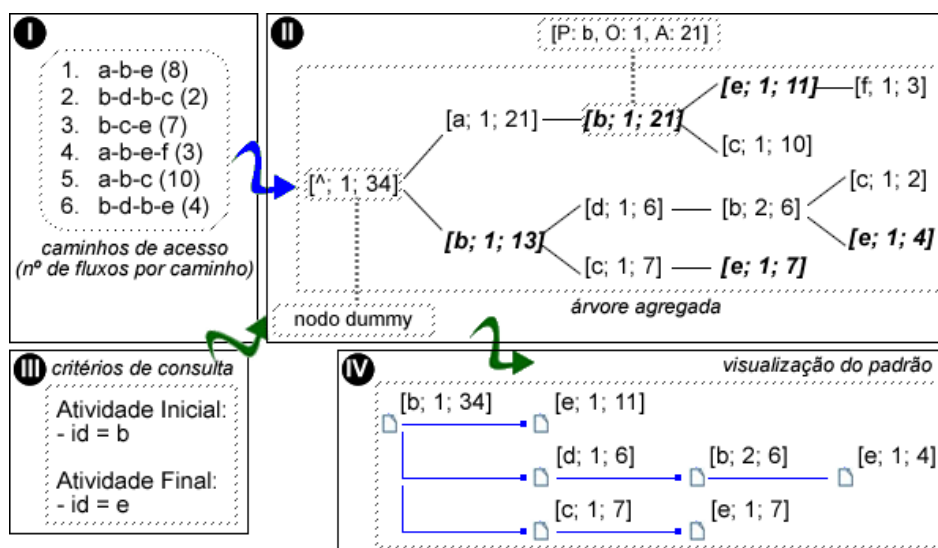


Figura 5 – Mineração e visualização de padrões de navegação.

Sobre a árvore agregada aplica-se o algoritmo de mineração seqüencial mediante critérios de consultas (e.g. caminhos que começam na página “b” e terminam na página “e”, como ilustrado na Figura 5.III), resultando em padrões de acessos (sub-caminhos da árvore),

que são visualizados graficamente também na forma de árvore (Figura 5.IV). Nela, os diferentes fluxos da árvore agregada que atendem às características desejadas são unificados.

As propriedades (critérios) que o usuário pode definir para dirigir a consulta de padrões sequenciais são: (a) conteúdo (URL ou título que identifica a página inicial e/ou final), (b) estrutura (tamanho do padrão) e (c) estatística (a ocorrência de uma página na trilha, ou uma medida relativa à frequência de acessos feitos a uma página). No caso da medida estatística de frequência de acesso, as opções são o número absoluto de acessos, o suporte (proporção de acessos sobre o total de sessões no *log*), ou o grau de convergência (razão dos acessos entre a página inicial e final de um padrão), este último somente disponível para a página final. A Figura 6 representa a interface de consulta disponível em *WebPath* para a definição das propriedades e busca de padrões.

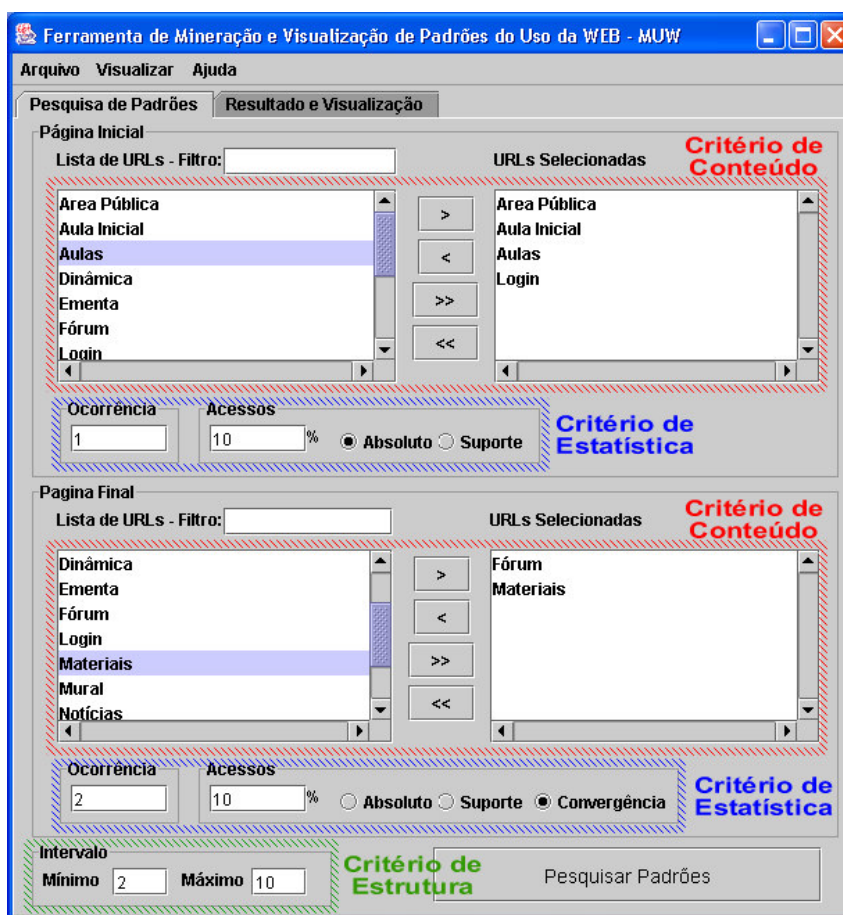


Figura 6 – WebPath: interface de consulta (definição de critérios)

No exemplo, o usuário solicita todos os padrões sequenciais que contenham como página inicial a página “Area de Pública”, “Aula Inicial”, “Aulas” ou “Login” e como página final a página “Fórum” ou “Materiais” (critério de conteúdo). Além disso, as páginas inicial e final devem possuir no mínimo 10 acessos, contudo a página inicial deve representar o seu

primeiro acesso na sessão e a página final o segundo (critérios de estatística). Outra restrição é com relação ao tamanho do padrão que deve conter entre duas e dez páginas (critério de estrutura).

A Figura 7 ilustra a interface de visualização (parte superior) e um padrão obtido com o uso de *WebPath* (parte inferior) na forma de árvore agregada.

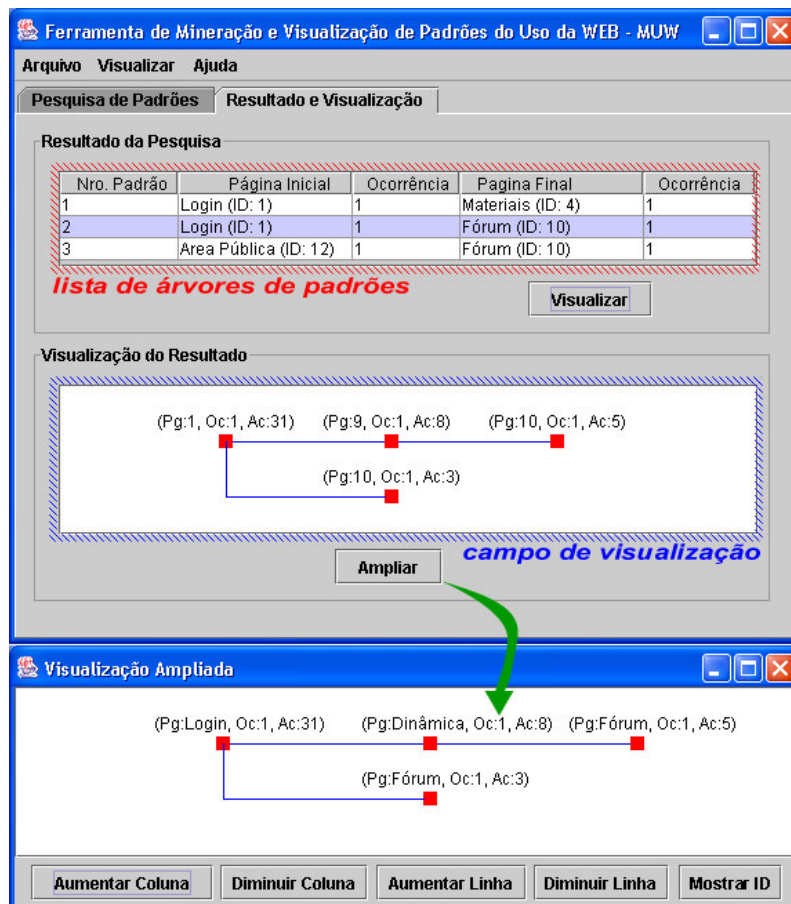


Figura 7 – WebPath: interface de visualização.

Analisando o exemplo, pode-se verificar que o padrão obtido descreve os possíveis caminhos para se chegar à página de Fórum, a partir da página de *Login* do site. Levando em consideração as informações de ocorrência e suporte, pode-se concluir que das 31 sessões que acessaram a página de *Login* pela primeira vez, ocorrência (Oc) igual à 1, apenas 8 chegaram na página de Fórum. Além disso, analisando as ramificações, pode-se observar que das 8 pessoas que chegaram na página de Fórum, 3 vieram diretamente da página *Login* e 5 passaram antes pela página de Dinâmica (Pg: 9). Maiores informações referente à interface de visualização podem ser obtidas consultando [TRI04].

2.4. Considerações

A gestão de processo de negócio possibilita um maior controle e acompanhamento das atividades pertencentes a um processo. Soluções para a gestão de PN devem levar em conta tanto a capacidade de automação e controle, quanto a capacidade de análise e monitoração dos dados de execução de PN.

Ter uma percepção do andamento dos negócios mais próximo da realidade possibilita identificar o quanto antes anomalias de execução dos processos, bem como prever possíveis falhas e simular melhorias tanto na definição quanto na execução de processos. Esse rápido poder de reação proporciona redução de custos, aumento da qualidade e satisfação dos clientes e maior valor da decisão tomada para o negócio [MEL04a].

A escolha e uso de uma ou de outra técnica de mineração depende muito do domínio de análise. Por exemplo, a técnica de classificação pode ser utilizada para gerar um modelo de previsão, e.g. árvores de decisão [GRI04]. Técnicas de mineração seqüencial têm sido utilizadas para descoberta do modelo do processo e alinhamento do negócio [AAL05]. Por outro lado, no domínio da mineração do uso da *web*, a técnica de mineração seqüencial tenta identificar as seqüências das páginas mais freqüentes, e assim, obter um melhor entendimento sobre o comportamento dos usuários, da estrutura e do conteúdo das páginas. Este benefício pode ser transposto à área de PN.

A seguir são apresentadas algumas abordagens dedicadas à redução da latência de dados, análise e decisão em PN. Para cada trabalho, verifica-se suas contribuições e limitações no que diz respeito à análise de PN. Desta forma, obtemos uma visão geral de todos os esforços na busca por descoberta de conhecimento em PN, evidenciando a inexistência de técnicas para a análise do fluxo de execução de atividades e a não integração de informação (falta de sinergia) entre as técnicas existentes.

3. TRABALHOS RELACIONADOS

Neste capítulo são apresentados alguns estudos relacionados à BPA, com a finalidade de expor o quanto já se evoluiu nesta linha de pesquisa. Para cada estudo são apresentadas técnicas para a solução dos problemas mencionados anteriormente, e sua devida contribuição para a redução do tempo entre a geração de um evento (execução de processos) e a tomada de decisão. Também é examinado se existe uma sinergia entre os componentes para análise, propostos pelas diferentes abordagens. Por fim, tendo em vista estes estudos e pesquisas para a obtenção de informação sobre dados de PN, tem-se uma visão geral do que é tratado ou não pelos trabalhos. Tendo em vista o cenário apresentado, realiza-se a caracterização do problema a ser resolvido por este trabalho.

3.1. Business Process Intelligence

Business Process Intelligence (BPI) é um ambiente proposto por Grigori et al. [GRI04] [SAY01], o qual possibilita a análise, monitoração, otimização, predição e controle de PN. Para suportar estas técnicas, BPI possui uma arquitetura muito similar àquela proposta por Golfarelli et al. [GOL04], descrita na Seção 2.2. BPI apresenta alternativas de solução para a redução das latências de dados, análise e decisão.

Para a redução da latência de dados, BPI conta com um extrator de dados de *logs* de processos (*PDW Loader*) e uma estrutura de armazenamento específica aos dados de processos de negócio, denominada *Process Data Warehouse* (PDW). O PDW [CAS02] é uma base analítica de dados composta por tabelas “fato” (objeto de análise) e “dimensão” (perspectiva de análise), organizadas estruturalmente pelo formato constelação de fatos [HAN01].

Desta forma, os dados podem ser analisados tanto por ferramentas comerciais, tais como *Crystal Reports* ou *Oracle Discoverer*, quanto por um componente específico para PN, denominado *Business Process Cockpit* (BPC). Adicionalmente, o PDW contém um grande número de informações agregadas que descrevem típicas métricas de desempenho, tal como a eficiência de um recurso. Desta forma, este modelo analítico possibilita, de forma rápida, a execução de diferentes tipos de análises e relatórios de acordo com as necessidades do analista. Por exemplo, usuários podem analisar quantas instâncias de processo, em um

intervalo de tempo, são iniciadas por certo usuário (análise do fato processo em relação às dimensões tempo e recurso).

Como solução ao desafio da redução da latência de análise, BPI dispõe de uma variedade de técnicas de mineração e recursos analíticos sobre os dados do PDW. Para permitir uma melhor interpretação e compreensão da execução dos processos, foi incorporada no PDW a análise semântica de processos, a qual objetiva a categorização das execuções de PN com relação a conceitos do domínio dos processos de negócio (e.g. “aprovado” vs “rejeitado”, “rápido” vs “médio” vs “lento”). A análise semântica de processos é apresentada em três formas: análise de comportamento, definição de taxonomia e regiões de processo.

A análise de comportamento permite a identificação de instâncias de processos que possuam características de interesse para o analista. Ela é realizada através de *templates*. Um *template* é uma definição parametrizada de um comportamento como, por exemplo, “instâncias de processo P nas quais um nodo N tenha sido executado mais do que T vezes”. Para realizar uma análise de comportamento o usuário deve apenas instanciar um destes *templates* fornecendo os valores para os parâmetros pré-definidos. Os *templates* são implementados por meio de consultas SQL.

Definição de taxonomia possibilita classificar instâncias de processo definidas pelo usuário. Cada taxonomia pode ser associada a um comportamento, significando que uma determinada instância de processo é classificada por certa taxonomia e se a instância de processo possui tal comportamento.

Região de processo é uma parte de um processo que pode ser tratada como uma única unidade de uma perspectiva de análise de negócio. Regiões de processo são definidas pela seleção de um nodo inicial i e um conjunto E de nodos finais f em um determinado processo P , o qual i e E identificam um sub-grafo G de P . Este sub-grafo deve satisfazer certas restrições:

- i tem somente um arco de entrada, vindo de fora de G ;
- somente nodos f não terminais podem possuir arcos com nodos de fora de G , sendo que este nodo f deve possuir apenas um arco de saída;
- todos os outros arcos de G devem ligar nodos contidos em G .

Pode-se relacionar região de processo com a definição de um sub-processo. A idéia é poder abstrair um conjunto de atividades contíguas, que não sejam de interesse da análise, em uma espécie de caixa preta, onde o analista tenha apenas a visão de entrada e saída desta região. Com o uso desse tipo de recurso busca-se facilitar a compreensão e análise da

execução de PN, eliminando nodos, ou conjuntos de atividades, de acordo com o interesse de análise dos diferentes níveis organizacionais.

No que tange às técnicas de mineração, BPI possui um componente específico, o *Process Mining Engine* (PME). Ele utiliza a técnica de classificação (árvores de decisão) tanto para a análise de comportamentos específicos, tal como a execução de certos caminhos em instâncias de processo, quanto para a geração de modelos de predição, usados para prever o comportamento e o desempenho de instâncias de processos, atividades e recursos.

Como solução ao desafio da redução da latência de decisão, BPI possui o componente *Business Process Cockpit* (BPC). Ele é uma camada de apresentação baseada em uma interface gráfica que provê relatórios, alertas de situações críticas e definição de consultas de modo intuitivo, sem a necessidade de escrever código ou conhecimento prévio por parte do usuário. Uma importante característica do BPC é a visualização dos dados de execução de processos (e de métricas de qualidade relacionadas) de acordo com diferentes perspectivas: sistema, processos, recursos ou serviços. Uma perspectiva identifica a entidade de processo que é o foco da análise. Por exemplo, sobre uma perspectiva de serviço, o usuário deve ser capaz de visualizar métricas sobre os serviços solicitados durante a execução do processo de negócio. No entanto, não há uma sinergia entre as diferentes formas de obtenção de informação.

3.1.1. iBOM

Castellanos et al. [CAS05b] apresentam uma plataforma para o gerenciamento de operações de negócio inteligente denominada iBOM. Essa plataforma pode ser considerada uma extensão da ferramenta BPI (Seção 3.1), a fim de solucionar algumas limitações presentes na primeira abordagem.

iBOM apresenta um novo componente denominado *Abstract Process Monitor* (APM), o qual incorpora a idéia de visão abstrata de processo. A partir deste componente, o usuário define quais atividades de um processo ele deseja analisar/monitorar desde a captura dos eventos. Desta forma, os usuários vêem os processos com diferentes níveis de abstração e diferentes visões de análise podem ser obtidas sobre o mesmo conjunto de dados. É importante ressaltar que, diferentemente da abordagem anterior (BPI), limitada a um sistema de *workflow*, APM possibilita a definição e monitoração de atividades executadas por sistemas heterogêneos, distribuídos e autônomos.

iBOM também permite que o usuário possa definir de forma gráfica métricas em diferentes perspectivas, bem como especificar a forma de computá-las. A grande contribuição de iBOM está em permitir ao usuário que ele possa escolher o que medir e de que forma isto será computado, dando a liberdade de medir o processo da forma que melhor lhe convier. Cada métrica é caracterizada por:

- Nome único;
- Tipo de objeto a ser medido (recurso, instância de processo, ou tipo de processo), também mencionado como foco. Define a abrangência da medição. Por exemplo, se for uma instância de processo a métrica é aplicada a cada instância;
- Consulta para a computação da métrica;
- Tipo de dado que representa o tipo de métrica, que pode ser numérico ou lógico;
- Indicações de quais valores de métricas são consideradas desejáveis ou indesejáveis (i.e. KPIs). Por exemplo, para métricas numéricas, isto é especificado pela definição de escalas com algum valor associado (verde, amarelo, vermelho).

Além das técnicas disponibilizadas por BPI como, por exemplo, árvore de decisão (classificação), iBOM incorpora uma nova técnica de otimização denominada *what-if analysis*. Nesta técnica, o usuário cria uma definição de processo abstrato e verifica o seu comportamento de execução. Depois, ele realiza modificações sobre esse modelo, inserindo novos valores aos parâmetros do processo abstrato, simulando configurações de execução para verificar o comportamento deste processo, caso apresente determinado cenário de execução. Com isso é possível comparar as diferentes situações de execução de um processo e conhecer quais cenários são propensos a apresentar resultados indesejáveis.

3.2. Delta Analysis e Conformance Testing

Van der Aalst [AAL05] afirma que mineração de processos objetiva prover técnicas e ferramentas para descobrir padrões de processos, controle e dados a partir de *logs* de eventos. A idéia básica da mineração de processos é diagnosticar processos de negócio pela mineração de eventos registrados em *logs* de execução. A mineração de processos proporciona três tipos de informação: descoberta, comparação e predição de processos. A descoberta objetiva a geração de um modelo de processo, quando não definido em um modelo explícito anteriormente. Dessa forma, pela identificação das atividades e seqüência em que são realmente executadas, é possível descobrir como as pessoas e procedimentos do negócio

interagem. Quando a organização tem um modelo pré-definido, a mineração de processos objetiva verificar o alinhamento do negócio através da comparação do modelo pré-definido com a real execução. Por exemplo, pode-se identificar que caminhos originalmente modelados não estão sendo executados ou que estão sendo executados caminhos não previstos no modelo original. A predição é usada para detectar, o quanto antes, comportamentos inesperados que requerem ações corretivas como, por exemplo, alterar alguma medida de desempenho quando utilizado uma técnica de mineração preditiva baseada em dados históricos (e.g. árvores de decisão [GRI04]).

Van der Aalst [AAL05] apresenta duas técnicas de mineração de processos relacionadas à comparação: *Delta Analysis* e *Conformance Testing*. *Delta Analysis* é utilizada para comparar o processo real, representado por um modelo de processo obtido através de mineração de seqüências de atividades, com algum processo pré-definido representando o sistema de informação. Contudo, esta abordagem tem dois problemas: (a) é necessário utilizar outra técnica de mineração para descobrir a partir de *logs* de execução o processo real e (b) ela não provê medidas quantitativas para medir a relação de similaridade entre o modelo definido e o gerado pelo *log*. *Conformance Testing* é abordada como uma solução alternativa a estes problemas. A idéia básica é comparar o modelo de processo pré-definido diretamente com o *log* de eventos, eliminando a necessidade de ter que descobrir o modelo real do processo. Adicionalmente, são consideradas algumas métricas para o teste de conformidade.

Em ambas as abordagens, a visualização dos modelos e análise dos resultados não é uma tarefa trivial, dada a inexistência de mecanismos para abstrair modelos. A Figura 8 ilustra a visualização de dois modelos de processos: o modelo pré-definido (à esquerda da Figura 8) e o real executado (à direita da Figura 8).

Além disso, só é possível analisar o processo como um todo, não dispondo de filtros ou recursos analíticos para a análise de comportamentos específicos de execução ou partes do processo que representem fluxos de execução de interesse (e.g. identificar caminhos que convergem para uma atividade específica ou encontrar os possíveis caminhos dada uma atividade inicial). A falta de informações associadas aos modelos, pré-definido e atualmente executado, não permite compreender as causas de discrepâncias entre eles.

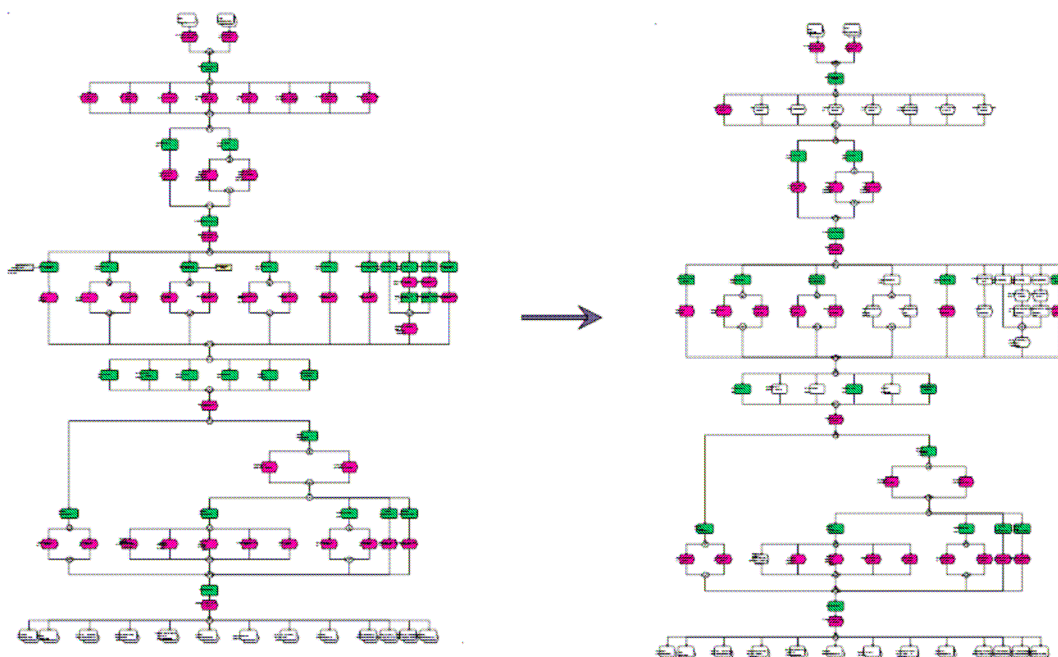


Figura 8 – Visualização de processos com Delta Analysis.

3.3. Corporate Performance Measurement System

List et al. [LIS04] propõem a idéia do *Corporate Performance Measurement System* (CPMS). A grande contribuição de um CPMS é a integração de dados do processo (*logs* de WfMS) com informações referentes ao negócio (*Corporate Data Warehouse* - CDW), provendo uma fonte única de dados genérica, com informações de desempenho em diferentes granularidades para o acesso dos diferentes níveis organizacionais. A partir desse modelo pode-se responder a uma série de questões (sumarizações) de negócio como, por exemplo, quantas reclamações são feitas por clientes especiais (medida pelo número de reclamações sob dados externos) ou qual é o tempo de trabalho de um processo ou atividade (medida pelo tempo de trabalho sob dados do *log*). Além disso, o CPMS possibilita diferentes tipos de análise de dados e visualização das informações.

3.4. Process Information Factory

Schiefer et al. [SCH04] propõem uma arquitetura, denominada *Process Information Factory* (PIF), como uma solução para o gerenciamento dos dados de processos de negócio dispersos entre diferentes aplicações. O foco principal da arquitetura é a

integração dos dados do negócio, presente nos sistemas de TI, com os dados e métricas de processo, disponíveis em DW tradicionais, em um *Enterprise Data Warehouse* (EDW), provendo uma análise mais completa da situação do negócio. Além disso, a partir dos dados contidos no EDW pode-se gerar relatórios de acordo com o interesse do usuário ou *Data Marts* contendo dados específicos para a análise dos diferentes setores da organização. PIF também possui um componente destinado à monitoração e cálculo de métricas de PN, denominado *Event Processing Container* (EPC).

3.5. MUW x BPA

Tristão et al. [TRI06] reporta um estudo onde *WebPath* [TRI04] foi utilizado no domínio de BPA. Esse estudo objetiva analisar o comportamento de instâncias de processos a partir do uso da técnica de mineração sequencial presente em *WebPath*, bem como apresentar os benefícios e dificuldades dessa prática. Neste sentido, pôde-se constatar que a técnica de mineração de seqüências pode ser benéfica na análise e visualização dos diferentes caminhos (seqüência de atividades) executados pelas instâncias de processos, permitindo compreender comportamentos específicos de execução ou identificar anomalias no fluxo.

Este estudo foi possível por haver uma similaridade de conceitos entre os dois domínios (MUW x BPA). O Quadro 1 apresenta esta similaridade de conceitos. Na MUW a navegação em um site corresponde a sessões que representam o acesso de usuários a uma seqüência de páginas em certa data e hora. Em BPA, a execução de um processo corresponde às instâncias de processo que representam a execução de uma seqüência de atividades por determinados recursos em certa data e hora.

Quadro 1 – Relação entre conceitos.

MUW	BPA
Navegação/acesso	Execução
Sessão	Instância de processo
Páginas	Atividades
Usuários	Recursos
Hora de acesso	Hora de execução

Como resultado deste estudo, pôde-se concluir que a técnica de mineração sequencial, quando aplicada no contexto de BPA, possibilita obter algumas informações, tais

como: (a) acompanhamento preciso da execução das atividades, (b) nível de convergência entre as atividades e (c) possíveis anomalias de execução do processo. Além disso, *WebPath* permite, por meio de consultas sobre a árvore agregada e especificando alguns critérios (conteúdo, estrutura e estatística), que o usuário possa definir (restringir) um menor número de padrões ou visualizar caminhos de execução específicos ou com características de acordo com o seu interesse.

Tristão et al. [TRI06] também apontam algumas limitações do *WebPath* para a análise e interpretação de dados de PN: (a) não suporta a representação de sub-processos, pois atividades compreendidas por um sub-processo são apresentadas como pertencentes ao processo principal (processos aplainados); (b) não fornece visões do processo segundo diferentes óticas da organização, i.e. não é possível selecionar um conjunto de atividades de interesse analítico; (c) não suporta a simplificação (generalização) ou detalhamento (especialização) dos processos; e (d) não revela métricas ou outras informações analíticas mais globais sobre o processo, apesar de revelar detalhes estruturais.

3.6. Considerações

O presente capítulo apresentou alguns estudos relacionados à BPA, os quais buscam alternativas e soluções para alguns dos desafios para a análise em tempo real, conforme apresentados na Seção 2.2. O Quadro 2 apresenta um resumo, em quatro aspectos, das propostas e contribuições de cada trabalho, que são: (a) a presença de um modelo para armazenar os dados de PN, (b) o tipo de análise realizada sobre os dados, (c) a possibilidade de troca de informações entre diferentes técnicas de análise, e (d) a disponibilidade de obter diferentes visões sobre o processo. Observa-se na Quadro 2 que todos os trabalhos possuem uma forma de armazenar os dados do processo, com uma pequena restrição à mineração do uso da *web* que armazena, apenas, informações de fluxo de execução para aplicar algum mecanismo de análise sobre os dados. Contudo, apenas a técnica de mineração de seqüência do uso da *web* possibilita analisar o fluxo de execução das instâncias de processo. Vale destacar que algumas abordagens até combinam diferentes técnicas de análise, mas de forma dissociada, não apresentando recursos para viabilizar a sinergia entre diferentes técnicas.

Quadro 2 – Quadro resumo dos trabalhos pesquisados.

Proposta \ Aspecto	Modelo de Dados	Análise de Dados	Sinergia de Análise	Abstração sobre o Processo
Business Process Intelligence	Process Data Warehouse (modelo analítico)	<ul style="list-style-type: none"> • Análise semântica – Técnica de classificação 	<i>Não trata</i>	Business Process Cockpit (OLAP e Alertas)
iBOM	Processo Abstrato (interface a diferentes fontes de dados)	<ul style="list-style-type: none"> • Métricas – Técnica de classificação – Técnica de otimização (<i>what if analysis</i>) 	<i>Não trata</i>	Abstract Process Monitor (processo abstrato - diferentes níveis de análise)
Delta Analysis e Conformance Testing	<i>Não trata</i>	<ul style="list-style-type: none"> • Técnica de seqüência (comparação de modelo e alinhamento do negócio) 	<i>Não trata</i>	<i>Não trata</i>
Corporate Performance Measurement System	Corporate Data Warehouse	<ul style="list-style-type: none"> • Sumarizações 	<i>Não trata</i>	<i>Não trata</i> (diretamente relacionada com a técnica de mineração aplicada)
Process Information Factory	Enterprise Data Warehouse	<ul style="list-style-type: none"> • Relatórios 	<i>Não trata</i>	Geração de Data Marts
Mineração Sequencial do Uso da Web	Árvore agregada (informações de fluxo de atividades)	<ul style="list-style-type: none"> • Mineração: técnica de seqüência (análise dos fluxos de execução) 	<i>Não trata</i>	Seleção de fluxos específicos de interesse

4. FLOWSPY

FlowSpy é um ambiente de apoio à análise de processos de negócio, o qual possui recursos para a análise de processos de negócio que combinam e permitem explorar de forma sinérgica as informações advindas da aplicação de técnicas de mineração seqüencial, originalmente propostas para a MUW, com aquelas sobre mensuração de processos, atividades e recursos através de KPIs.

FlowSpy utiliza a técnica de mineração seqüencial proposta por Spiliopoulou et al. [SPI00] [SPI98] por possibilitar uma análise exploratória dos diferentes fluxos de execução de atividades através da realização de consultas que permitam expressar critérios representativos sobre padrões de execução. Obtém-se como resultado caminhos específicos a comportamentos de execução, voltados às reais necessidades do usuário. *FlowSpy* incorpora, à técnica de mineração seqüencial do uso da *web*, mecanismos de abstração para auxiliar o analista tanto na seleção de atividades de interesse presentes no modelo, reduzindo o domínio de pesquisa, quanto na visualização e interpretação dos resultados, generalizando ou especializando o modelo, apoiando diferentes visões do negócio e simplificando a análise dos resultados.

Com o uso dos mecanismos de abstração é possível: (a) definir perfis de análise, os quais permitem ao usuário definir/organizar atividades e/ou sub-processos (conjunto de atividades) em um grupo de análise; (b) filtrar o *log* de entrada, reduzindo o número de atividades a serem processadas, excluindo atividades ou perfis de análise que não sejam de interesse do analista; e (c) visualizar padrões de navegação abstraindo atividades, sub-processos e perfis de análise, possibilitando aumentar ou diminuir os níveis de detalhamento dos padrões resultantes (análogo às operações *drill-up* e *drill-down* de OLAP).

4.1. Arquitetura do Ambiente

Análise de processos de negócio (BPA) é um campo de pesquisa amplo, que compreende desde a captura e preparação dos dados, até a aplicação de técnicas de análise e visualização dos mesmos, como refletido na arquitetura proposta por Golfarelli et al. em [GOL04], apresentada na Seção 2.2.

O foco do ambiente proposto é explorar a sinergia entre duas técnicas de análise de processos: a análise estrutural, pelo emprego da técnica de seqüência da MUW, e análise

quantitativa, pelo emprego de KPIs. A arquitetura do ambiente proposto reflete este propósito, não levando em consideração a forma de captura dos dados, estruturas específicas de armazenamento e outras técnicas de análise. Posteriormente, os demais componentes da arquitetura de referência da Golfarelli [GOL04] podem vir a ser integradas nesta proposta. A Figura 9 apresenta esta arquitetura de ambiente, a qual propicia a sinergia entre a análise estrutural (mineração sequencial) do processo e dos recursos analíticos baseados em KPIs.

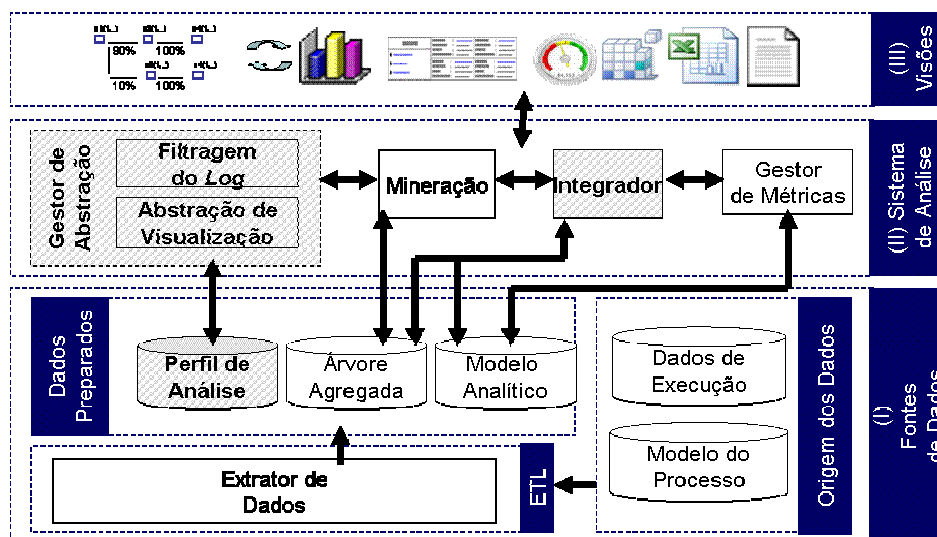


Figura 9 – Arquitetura do ambiente FlowSpy.

A arquitetura está dividida em 3 camadas, conforme ilustra a Figura 9. A camada (I) fonte de dados armazena e organiza os dados de definição e execução de processos em um único repositório de dados. Na camada (II) sistema de análise, diferentes componentes atuam sobre esses dados e permitem a definição de metadados, métricas e indicadores de desempenho de processos, e a aplicação de técnicas de mineração. A camada (III) visão apresenta diferentes formas de visualização desses dados de acordo com as necessidades da organização. Os componentes representados de forma hachurada na Figura 9 (i.e. Perfil de Análise, Gestor de Abstração e Integrador) são desenvolvidos e introduzidos por essa proposta, e constituem o cerne de sua contribuição. Os diferentes componentes da arquitetura são descritos em maiores detalhes no restante desta Seção.

4.1.1. Origem dos Dados

Diferentes sistemas podem estar sendo utilizados para gerenciar e controlar os processos de negócio e, com isso, diferentes fontes de dados podem armazenar informações desses processos. A lógica do negócio pode estar fragmentada nesses diferentes ambientes

heterogêneos. Outros sistemas podem ainda conter informações para contextualizar o processo, e.g. dados organizacionais.

Para essa proposta, é indiferente a origem dos dados. A única restrição é que, dessa fonte de dados, possa se obter o conjunto ordenado de atividades executadas por cada instância de processo como, por exemplo, *logs* de execução, representado na Figura 9 pelo componente Dados de Execução. Este tipo de informação é essencial para a realização da análise seqüencial da execução de processos.

O componente Modelo do Processo possui informações relacionadas à definições de processos, atividades e recursos, as quais são utilizadas para a geração de métricas e contextualização de sumarizações.

4.1.2. Modelo Analítico

Uma solução para tratar a heterogeneidade de sistemas é armazenar os dados em uma única fonte de dados e, assim, facilitar a tarefa de mineração e descoberta de conhecimento. Trabalhos como [LIS04], [SCH04] e [GRI04] discutem formas de capturar e armazenar dados do processo.

O Modelo Analítico é uma representação de uma estrutura de armazenamento de dados, o qual sua organização facilita o processo de descoberta de conhecimento. O ponto positivo em se utilizar um modelo analítico é permitir que a arquitetura possa ser livre quanto ao WfMS utilizado, pois ele independe da estrutura de armazenamento fornecida pelo WfMS.

Nessa proposta, não importa o tipo de Modelo Analítico utilizado para armazenar os dados do processo. Neste caso, a riqueza dos dados influencia apenas o tipo e qualidade das métricas de análise.

4.1.3. Extrator de Dados

O extrator de dados é responsável pela tarefa de Extração (*Extraction*), Transformação (*Transformation*) e Carga (*Loading*) (ETL) dos dados existentes nas fontes de dados para os dados preparados. Na abordagem *FlowSpy*, o extrator de dados, além de ser responsável pela carga do modelo analítico, também é responsável pela geração do *log* de execução utilizado para a carga/construção da estrutura de dados de armazenamento dos diferentes fluxos, denominada árvore agregada (Seção 2.3.1).

4.1.4. Perfil de Análise

Perfil de análise é a denominação de um recurso criado para representar um conjunto de atividades que definem o interesse particular de análise de uma pessoa ou grupo. Perfil de análise é similar à abordagem de região de processo [GRI04], descrito na Seção 3.1, que permite que um conjunto de atividades possa ser tratada como uma única unidade de execução.

Os perfis de análise podem ser definidos em termos de: atividades e sub-fluxos de processos. Uma atividade é uma unidade de execução de um processo P . Um sub-fluxo de processo é definido como um grafo dirigido $SG = \langle N, L \rangle$, composto de um conjunto de nodos N e ligações L , onde $s, s \in N$, é o nodo inicial, $F, F \subset N$, é o conjunto de nodos finais, e $L \subset N \times N$ é o conjunto de ligações definidas entre nodos.

Nota-se que um sub-processo é um tipo de sub-fluxo. Desta forma, se o modelo analítico de dados possuir informações do modelo do processo, a informação de sub-processos pode ser utilizada para a geração de perfis de análise. A Figura 10 ilustra a criação de perfis de análise (Figura 10.II) a partir do modelo de um processo de negócio (Figura 10.I). O perfil de análise PA1 representa as atividades do sub-processo S1 e o perfil PA2 as atividades do sub-processo S2, enquanto o perfil PA0 representa as atividades e sub-processos (perfis PA1 e PA2) do processo principal.

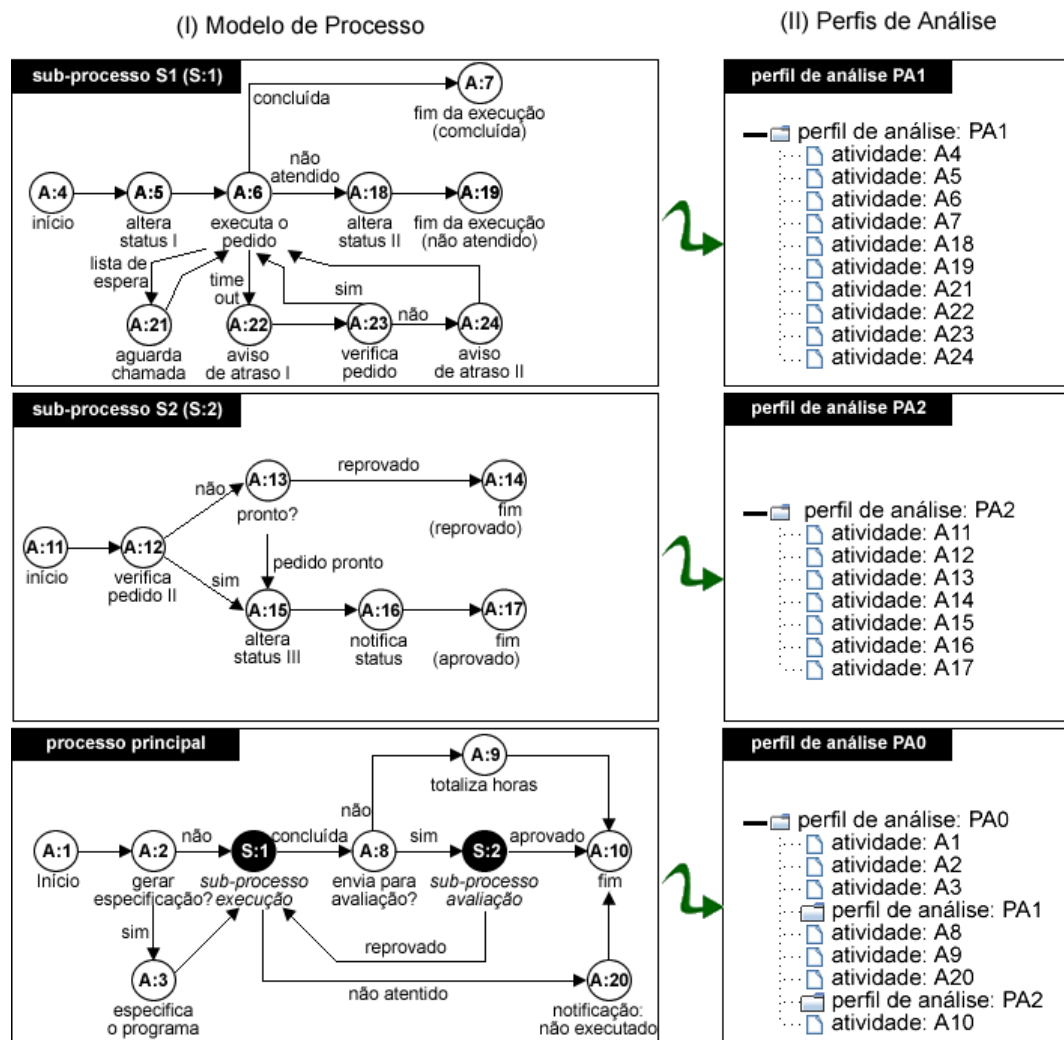


Figura 10 – Criação de perfis de análise a partir do modelo de um PN.

Para criar um perfil de análise, o usuário deve selecionar um nodo inicial e, interativamente, *FlowSpy* mostra os nodos seguintes (i.e. a partir do *log* de execução, os nodos que podem ser imediatamente executados a partir dele) que podem ser selecionados pelo usuário. Isto pode ser feito até que seja(m) definido(s) nodo(s) final(ais) do sub-fluxo. Caso contrário, se o modelo do processo estiver disponível, a interface mostra ao usuário que já existem sub-processos definidos e a estrutura do processo é usada para indicar os nodos seguintes. Uma vez que o perfil esteja definido, o gestor de abstração pode utilizá-lo para a filtragem do *log* de criação/geração da árvore agregada ou durante a abstração de visualização.

4.1.5. Gestor de Abstração

Gestor de abstração é o componente de *FlowSpy* responsável pelo gerenciamento dos recursos de abstração de processos, composto por: (a) filtragem do *log* e (b) visualização de padrões.

4.1.5.1. Filtragem do Log

Filtragem do *log* objetiva a geração de uma árvore agregada mais compacta, contendo somente as atividades/sub-fluxos de interesse como, por exemplo, as representadas por um dado perfil de análise. O *log* de execução pode ser filtrado de forma inclusiva ou exclusiva. Na forma inclusiva, são carregadas apenas as instâncias de atividades selecionadas. Se as atividades/perfis de análise forem utilizados de forma exclusiva, a filtragem implica em remover do *log* todas as atividades e sub-fluxos selecionados. Uma outra possibilidade da filtragem do *log* na forma exclusiva é agregar as atividades selecionadas em um nodo único de execução. A Figura 11 ilustra um exemplo de filtragem do *log* nas formas inclusiva e exclusiva (agregação).

Seja o processo modelado na Figura 11.I e a árvore agregada (Figura 11.II) originada pela carga completa do *log* de execução, o objetivo é verificar o comportamento de execução desse processo desconsiderando as atividades 6, 7, 8, 9, 10 e 11, pelo uso do perfil de análise PA3 contendo estas atividades. A Figura 11.III mostra a simplificação da árvore agregada resultante após a filtragem exclusiva deste *log* utilizando o perfil de análise PA3. Nota-se que as atividades da árvore resultante, pertencentes ao perfil selecionado, são substituídas por nodos P3. Já, a Figura 11.IV ilustra um exemplo de filtragem do *log* de execução na forma inclusiva, onde a árvore resultante apresenta apenas as atividades pertencentes ao perfil de análise PA3.

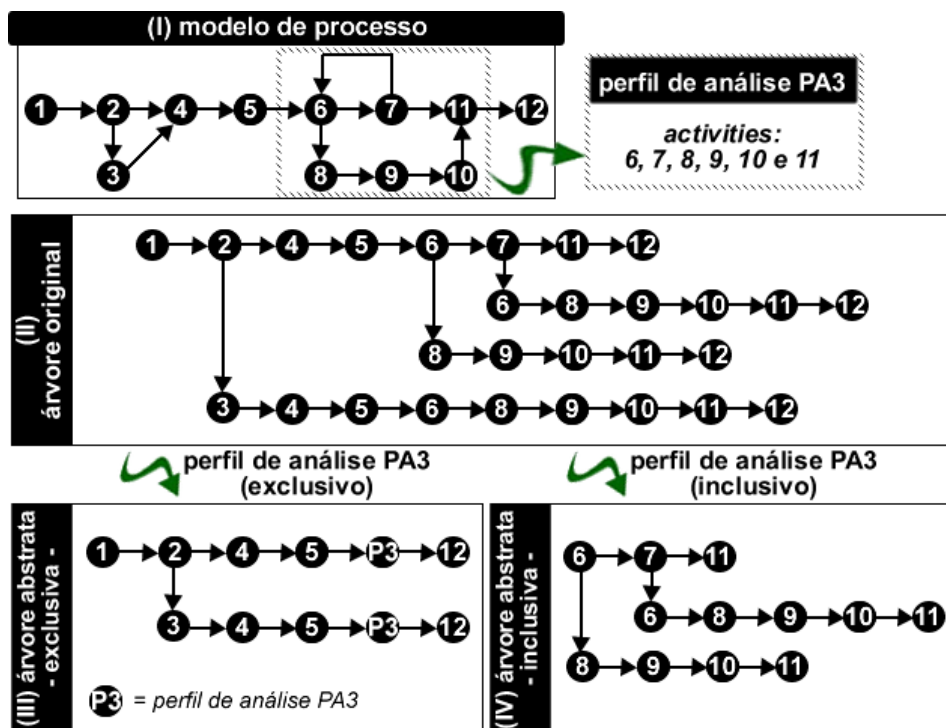


Figura 11 – Filtragem do log de execução

Como consequência da abstração provida pela filtragem do *log*, a tarefa de análise é simplificada e, conseqüentemente, torna-se mais eficiente pelo fato da árvore se tornar significativamente menor. Dessa forma, o usuário pode escolher um conjunto simplificado de atividades para definir as consultas de mineração e, assim, o padrão resultante irá listar um número consideravelmente menor de atividades.

4.1.5.2. Abstração de Visualização

Abstração de visualização é o recurso de abstração que possibilita ter diferentes visões de um mesmo padrão de execução pela sua simplificação. A idéia é análoga às operações de *drill-up* e *drill-down*, normalmente usadas por ferramentas OLAP. Desta maneira, o usuário pode aumentar ou diminuir o nível de detalhe do fluxo representado pelo padrão de acordo com seu interesse. A simplificação pode ser baseada em perfis de análise, ou interativamente pela seleção das atividades existentes no padrão. Na forma interativa, leva-se em consideração as mesmas restrições para a definição de um perfil de análise que represente um sub-fluxo. A visualização de padrões pode ser feita de duas formas: *agregação* e *exclusão*.

Na *agregação*, o resultado é um padrão de execução simplificado pela substituição das atividades e perfis de análise por um nodo atômico. A Figura 12 mostra um

exemplo de abstração de visualização do tipo agregação. Nesse exemplo, um padrão obtido a partir do modelo de processo utilizado como exemplo na Seção 3.5 é simplificado pela aplicação do perfil de análise PA1. Desta forma, todas as atividades relacionadas ao sub-processo S1 (A:4, A:5, A:6, A:7, ... , A:18, A:19) são agrupadas em um único nodo ([P1, O, I]), facilitando sua interpretação.

Na *exclusão*, o usuário interage indicando iterativamente os nodos da árvore que devem ser removidos da visualização. Estes nodos podem corresponder a nodos atômicos ou sub-fluxos abstraídos anteriormente (modo agregação). A Figura 13 ilustra um exemplo de abstração de visualização do tipo exclusão, tomando como base o padrão resultante da Figura 12, sobre o qual foi aplicado a abstração de visualização do tipo agregação.

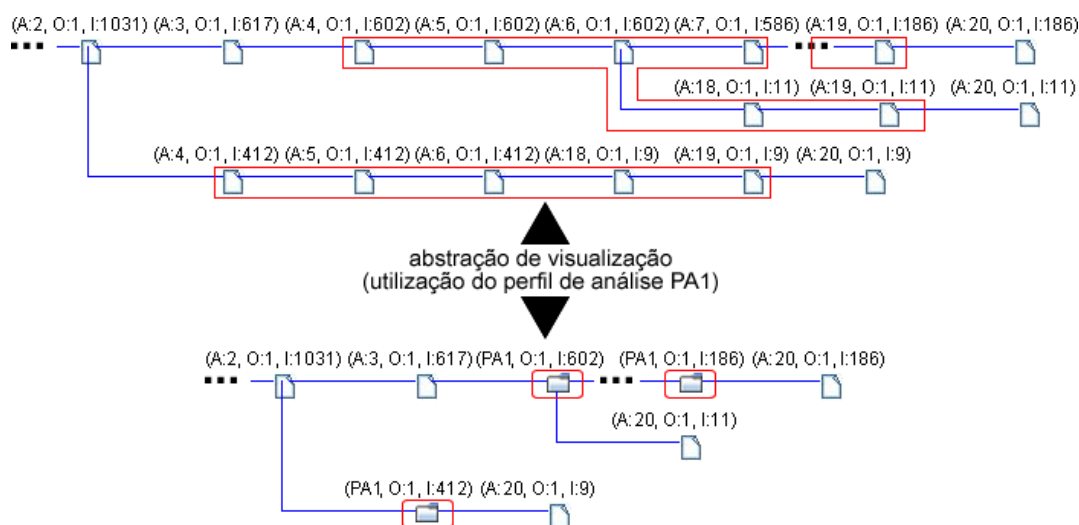


Figura 12 – Abstração de visualização do tipo agregação.

Por exemplo, considerando o padrão resultante da Figura 12 (Figura 13.I) e a hipótese de que o usuário deseja remover a atividade 3 (A:3). O algoritmo responsável pela exclusão das atividades ou perfis de análise apenas remove as atividades solicitadas, conforme ilustrado pela Figura 13.II. Após a exclusão do nodo, o algoritmo responsável pela abstração de visualização prevê a formação de caminhos semelhantes (e.g. na Figura 13.II existem dois caminhos iniciando pelo nodo [PA1, O: 1, A: _]). Desta forma, os caminhos são unificados e seus acessos somados, conforme padrão ilustrado pela Figura 13.III.

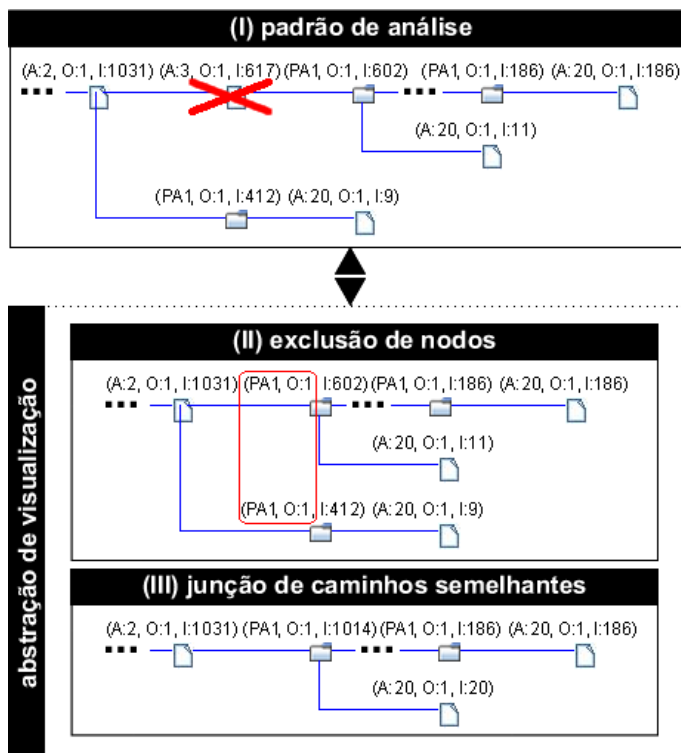


Figura 13 – Abstração de visualização do tipo exclusão

Na abstração de visualização de padrões é possível reverter as operações de abstração *agregação* e *exclusão* (i.e. expandir a visualização de nodos agregados ou re-incluir na visualização do padrão os nodos abstraídos). Esta é a principal diferença em relação às operações de filtragem, onde uma vez realizadas, não são possíveis de serem revertidas.

4.1.6. Mineração

Este componente corresponde à descoberta de padrões de comportamento a partir da investigação de dados históricos. Técnicas de classificação [GRI04] [SAY01] [CAS05b], utilizadas para análise, predição e simulação, e técnicas de seqüência [AAL05], para descoberta do modelo e alinhamento do PN, são alguns exemplos de técnicas de mineração utilizadas.

Na abordagem *FlowSpy*, foi utilizada a técnica de mineração seqüencial do uso da *web* [SPI00] [SPI98], por possibilitar o acompanhamento preciso da execução das atividades, a identificação de possíveis anomalias de execução e o nível de convergência entre as atividades (probabilidade de uma atividade executar um ou outro fluxo). Para tanto, foram

introduzidos alguns recursos para a melhor interpretação e análise dos padrões minerados (Seção 4.1.5).

4.1.7. Gestor de Métricas

O gestor de métricas é responsável pela análise qualitativa dos processos de negócio. Seu objetivo é realizar sumarizações sobre os dados de modo a medir o desempenho de execução dos PN, e verificar o comportamento destes mediante objetivos estabelecidos pela organização. As métricas são obtidas através de consultas que retornam valores sumarizados, e.g. o número de processos que encontram-se atrasados ou qual a média de tempo de execução de uma atividade.

4.1.8. Integrador

O componente integrador é responsável por propiciar a sinergia de informações entre a técnica de mineração seqüência do uso da *web* e análise quantitativa dos dados dos PN. Dessa forma, a partir de um fluxo de execução de processo, pode-se verificar o desempenho das instâncias de processos que o representam mediante indicadores que reflitam os objetivos da empresa. De fato, dado um conjunto de instâncias de processos que apresentam um desempenho para uma determinada métrica, pode-se tentar compreender o seu comportamento pela análise do fluxo de suas instâncias. A idéia de integração de informação das diferentes formas de análise de processos de negócio é resumida pela Figura 14.

Para possibilitar esta integração, a estrutura de dados, que representa a árvore agregada (Figura 5. II) e o padrão de execução das instâncias de processos (Figura 5. IV), deve conter, em cada nodo de execução, o conjunto de instâncias que passaram por este fluxo. A partir da seleção de um nodo (Figura 14.I), pode-se analisar informações quantitativas das instâncias de processos que a representam como, por exemplo, as instâncias de processo 01 e 02 do nodo E.

Um KPI possui associado um conjunto de instâncias de processos que ocasionam este desempenho. Desta forma ao selecionar um KPI (Figura 14.II) as instâncias relacionadas a este desempenho (no exemplo, as instâncias de processos 02, 04 e 06) são submetidas à análise estrutural e verificadas possíveis causas deste comportamento.

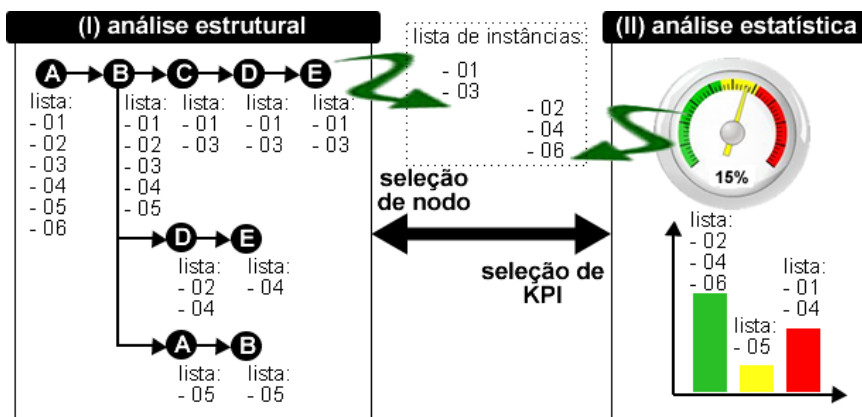


Figura 14 – Integração de informação.

A grande dificuldade da técnica de mineração seqüencial é identificar, em cada nodo, quais instâncias de processos executaram determinada atividade do fluxo. A solução é relacionar em cada nodo (atividade no fluxo) do padrão o conjunto de instâncias de processos que a executaram. Dessa forma, para cada nodo do padrão de execução tem-se o conjunto de instâncias de processos que seguiram determinado fluxo, sendo possível selecionar as instâncias e analisá-las quantitativamente, de acordo com o objetivo e indicador desejado.

A análise estrutural realizada a partir da análise quantitativa segue a mesma idéia. Ao selecionar um indicador de desempenho, é gerada a estrutura seqüencial contendo apenas as instâncias que ocasionam tal desempenho.

4.1.9. Visões

As visões são as portas de comunicação entre o ambiente *FlowSpy* e o usuário. Elas constituem desde interfaces de definição e configuração do sistema, até as diferentes formas de visualização dos dados, permitindo que um mesmo conjunto de dados possa ser analisado de diferentes formas, dependendo da disponibilidade e do interesse de análise. A forma de visualização dos dados está diretamente relacionada à técnica de mineração utilizada para análise de dados.

4.2. Considerações

Este capítulo apresentou a proposta de um ambiente de apoio à BPA denominado *FlowSpy*. A principal contribuição dessa abordagem, com relação às outras propostas de BPA (Capítulo 3), é permitir a troca de informação entre diferentes técnicas de análise e

mensuração de PN. Para isso, FlowSpy possui recursos para a análise de processos de negócio, que combinam e permitem explorar de forma sinérgica as informações advindas da aplicação de uma técnica de mineração seqüencial com aquelas sobre mensuração de processos, atividades e recursos através de KPIs.

FlowSpy enfoca apenas os componentes diretamente necessários à mineração e análise quantitativa de PN. Desta forma, a proposta desse ambiente independe da fonte de dados, do modelo utilizado para armazenar os dados de PN, do procedimento de ETL e do mecanismo para medir os processos. As visões do processo são diretamente ligadas às técnicas de mineração e análise, e aos recursos disponíveis para representar os dados do processo de negócio.

Em especial, esse ambiente possui o componente “Integrador”, que é responsável pela sinergia de informação entre as técnicas de análise abordadas (estrutural e quantitativa), e os componentes “Gestor de Abstração” e “Perfil de Análise”, que são responsáveis pela otimização na visualização de padrões de execução provenientes da técnica de mineração seqüencial do uso da *web*.

Fazendo uma análise das funcionalidades abordadas e propostas em FlowSpy e verificando os conceitos observados no Quadro 2, os quais resumem os trabalhos relacionados apresentados no Capítulo 3, pode-se chegar às seguintes conclusões:

- Modelo de dados – propõe-se a utilização de um modelo analítico para que os dados analisados independam da estrutura de armazenamento fornecida pelo WfMS.
- Análise de dados – pode-se realizar a análise quantitativa, pelo uso de métricas e KPIs, e análise estrutural, por meio da técnica de mineração seqüencial da MUW.
- Sinergia de análise – FlowSpy possibilita a troca de informação ente os dois modos de análise (estrutural x quantitativa).
- Abstração sobre o processo – os recursos de abstração pré-processamento do *log* e abstração de visualização possibilitam a abstração sobre o processo proporcionando diferentes formas de análise e visualização do processo.

5. EXPERIMENTAÇÃO

Neste capítulo são apresentados os casos de uso que envolvem ações sobre os componentes desenvolvidos neste trabalho: (a) perfil de análise, (b) gestor de abstração e (c) integrador, contemplados pelo ambiente de BPA proposto. Para tanto, no sentido de avaliar o ambiente, houve um esforço para prototipar estes componentes.

Nos experimentos realizados, por critério de simplificação, utiliza-se como fonte de dados a base de um WfMS (Oracle Workflow), já que ele possui tanto informações do modelo quanto de execução de processos e, além do mais, sua estrutura de armazenamento é de conhecimento do autor.

Por se trabalhar exclusivamente com dados de processos localizados em bases Oracle Workflow e por possuir diversas publicações e ser amplamente aceita no meio acadêmico, adotou-se, nos experimentos realizados, uma adaptação do modelo analítico apresentado por Casati et al. [CAS02] para armazenar os dados.

O extrator de dados é representado por um conjunto de scripts desenvolvidos em PL/SQL por Garcia et al. [GAR05], o qual extrai e transforma dados localizados em bases Oracle Workflow e os carrega para o modelo analítico proposto por Casati et al. [CAS02]. Para a construção/carga da árvore agregada foi utilizado um componente próprio da ferramenta WebPath, o qual monta a árvore a partir de um log de execução contendo um conjunto ordenado de atividades agrupadas e identificadas pelo identificador da instância de processo.

Como técnica de mineração, foi utilizada a técnica de mineração seqüencial do uso da *web* proposta por Spiliopoulou et al. [SPI00] [SPI98]. Para a análise quantitativa, utilizou-se a proposta de Castellanos et al. [CAS05b] presente na plataforma iBOM.

Os dados utilizados neste experimento representam 1031 instâncias de execução de um processo de negócio modelado na ferramenta *Oracle Workflow*. O processo analisado é um *workflow* real de solicitações de desenvolvimento de software, envolvendo 24 atividades, distribuídas em um processo principal e 2 sub-processos.

Para uma melhor compreensão do funcionamento dos componentes propostos no ambiente, cada caso de uso é comentado, trazendo descrições de como as operações são realizadas.

5.1. Modelo de Dados Simplificado

Alguns casos de uso necessitam exemplificar o acesso ao modelo de dados de PN. Dada a complexidade do modelo analítico proposto por Casati et al. [CAS02], é representada pela Figura 15, na forma de um diagrama de classes UML, um modelo de dados simplificado para que haja uma melhor compreensão e entendimento das consultas realizadas sobre o modelo analítico original. O modelo de dados simplificado é composto por cinco classes, sendo três do tipo dimensão (*Dim_KPI*, *Dim_Resource* e *Dim_Time*) e duas do tipo fato (*Fact_Instance_Process* e *Fact_Instance_Activity*).

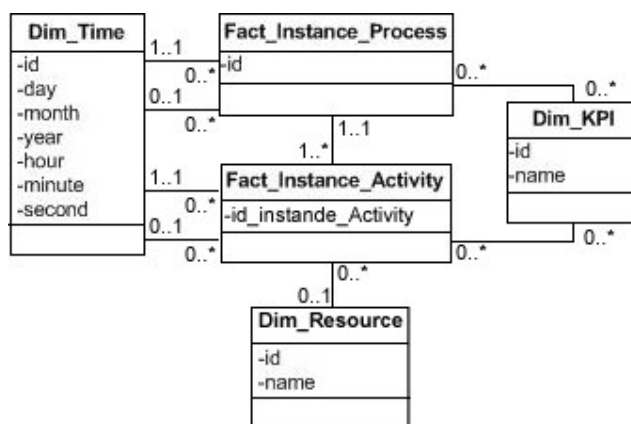


Figura 15 – Diagrama de classes simplificado.

Cada instância de processo em execução, ou já completada, fica armazenada como fato em *Fact_Instance_Process*. As atividades executadas por cada instância de processo são armazenadas em *Fact_Instance_Activity*. *Dim_Time* armazena o tempo em que a instância de processo e atividade iniciaram e terminaram sua execução, e *Dim_Resource* possui informações dos recursos que executam as instâncias de atividades.

As métricas em BPA são medidas que extraem algum tipo de informação, a partir de dados quantitativos, para que o gestor consiga verificar o andamento do processo e tomar decisões atendendo aos objetivos (estratégia) da organização, alterando e/ou otimizando processos existentes na empresa. Para essas métricas podem ser definidas indicadores de desempenho (KPIs) que dão significado aos valores obtidos pelas métrica. Originalmente, tanto a métrica quanto os KPIs correspondentes são armazenados em meta-modelos. No entanto, para simplificar a compreensão, os KPIs das métricas de instâncias de processos ou atividades são armazenados na dimensão *Dim_KPI*. Desta forma, um KPI pode representar nenhum ou várias instâncias e uma instância pode conter nenhuma ou vários KPIs de diferentes métricas.

5.2. Casos de Usos

Os casos de uso considerados pela interação do usuário com o protótipo envolvendo os recursos propostos no ambiente são apresentados na Figura 16, através de um diagrama de caso de uso de notação UML, os quais são detalhados abaixo.

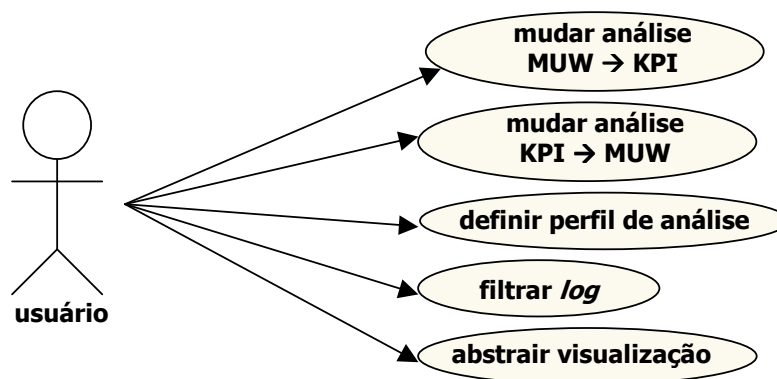


Figura 16 – Diagrama de casos de uso.

- *Mudar análise MUW → KPI*: após realizar a mineração seqüencial do uso da *web* e analisar um padrão de execução, o usuário seleciona um fluxo e obtém a análise quantitativa do conjunto de instâncias de processos associados ao fluxo escolhido.
- *Mudar análise KPI → MUW*: após realizar a análise quantitativa de um processo mediante uma métrica e analisar sua distribuição entre os indicadores definidos, o usuário seleciona um destes indicadores e obtém os fluxos de execução (análise estrutural) das instâncias de processo enquadradas em determinado desempenho.
- *Definir perfil de análise*: O usuário define um conjunto de atividades que representam o seu interesse de análise. Este perfil pode ser usado posteriormente pelo gestor de abstração.
- *Filtrar log*: o usuário simplifica o *log* para a geração da árvore agregada (Seção 2.3.1) selecionando o conjunto de atividades que não sejam de seu interesse analítico.
- *Abstrair visualização*: ao analisar um padrão gerado pela técnica de mineração seqüencial do uso da *web* (árvore de execução), o usuário seleciona um conjunto de atividades ou um perfil de análise para generalizar ou especializar o padrão resultante.

5.2.1. Caso de Uso Mudar Análise MUW → KPI

Este caso de uso representa a mudança da análise estrutural para a análise quantitativa. A interface do protótipo que realiza este caso de uso é mostrada na Figura 17. Destaca-se que para originar esse cenário, o usuário deve ter realizado uma consulta sobre a árvore agregada. A Seção 2.3.1 mostra um exemplo de consulta utilizando *WebPath*. O padrão de execução é visualizado no campo *Result Visualize* e é representado na forma de árvore agregada (Seção 2.3.1). Cada nodo desta árvore é representado por uma tripla [A, O, I], onde A representa a atividade executada, O a ocorrência desta atividade no fluxo e I o número de instâncias que executaram esta ocorrência de atividade.

Ao visualizar um padrão de fluxo de atividades, o usuário pode realizar a análise quantitativa de todo o padrão (pressionando o botão *Quantitative Analysis*) ou de uma parte dele, selecionando uma determinada ocorrência de atividade no fluxo, e.g. o nodo (A:3; O:1; I:617) circunscrito na Figura 17. O conjunto de 617 instâncias de processos que executaram esta ocorrência de atividade no fluxo é então analisado quantitativamente (Figura 18). Na tela de análise quantitativa, o usuário seleciona de um conjunto pré-definido de métricas, aquela que melhor responda as suas dúvidas. A Figura 18 ilustra a seleção da métrica *Count of Process Instances by Taxonomy*. Taxonomia é um conceito utilizado por Grigori et al. [GRI04] que possibilita classificar instâncias de processo definidas pelo usuário. Nesse exemplo, de acordo com o domínio do negócio, *Taxonomy* representa o tempo que determinada instância leva para completar seu fluxo de execução. As instâncias são classificadas em (a) *acceptable*, (b) *fast* e (c) *slow*.

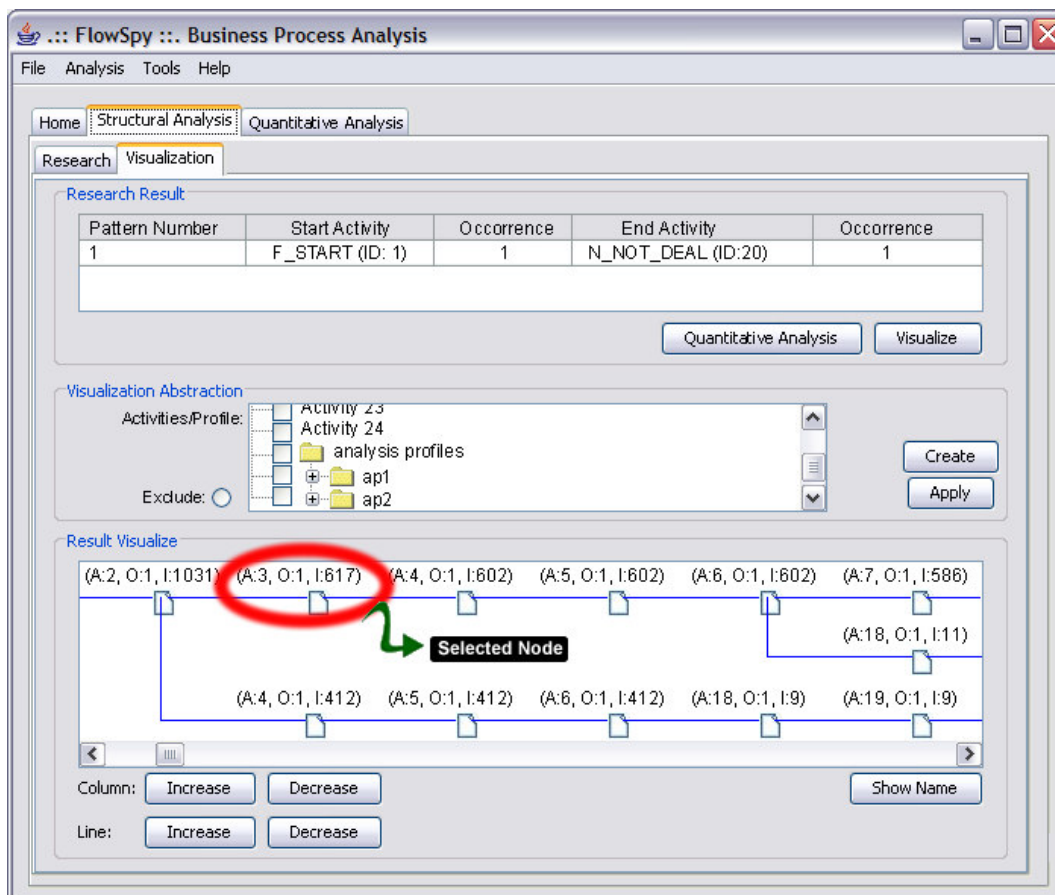


Figura 17 – Análise estrutural MUW → KPI.

Este tipo de análise possibilita verificar mediante um KPI a distribuição de comportamento daquelas 617 instâncias de processos que fazem parte do nodo de execução selecionado. Analisando a Figura 18, verifica-se que a grande maioria das instâncias (374) apresenta comportamento aceitável, outras 96 instâncias possuem uma execução rápida, e o restante (147 instâncias) demoram a completar todo o seu fluxo.

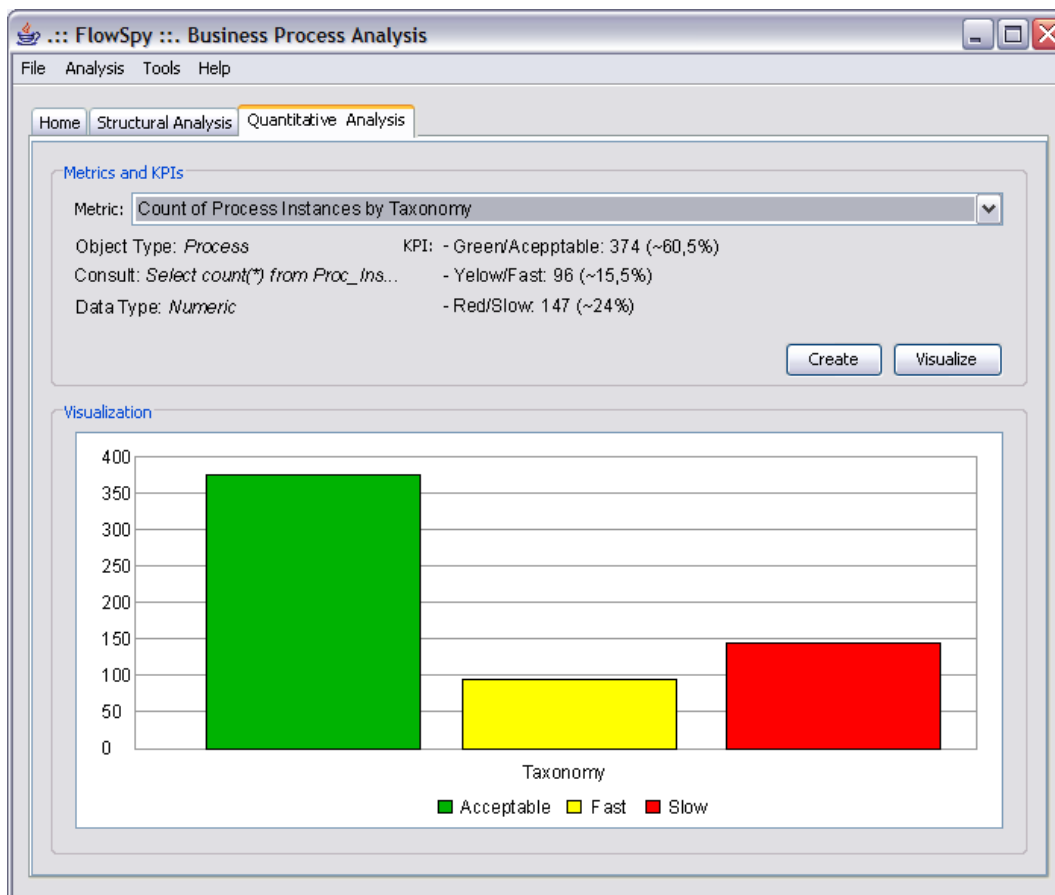


Figura 18 – Análise quantitativa MUW→KPI.

Para a geração do gráfico (ilustrado na parte inferior da Figura 18) que representa este comportamento de desempenho é realizada a consulta SQL ilustrada pela Figura 19.

SQL: Análise Estatística

1. Select Dim_KPI.name, count(Dim_KPI.id)
2. From Dim_KPI inner join Fact_Instance_Process on
Fact_Instance_Process.id_KPI = Dim_KPI.id inner join
Fact_Instance_Activity on Fact_Instance_Activity.id_instance_process =
Fact_Instance_Process.id
3. Where Fact_Instance_Process.id in (a₁, a₂, a₃, ..., a_n)
4. Group By Dim_KPI.name;

Figura 19 – SQL para métrica da Figura 17.

É importante destacar que o uso da cláusula *where* na consulta SQL restringe a consulta apenas sobre às instâncias de processo presentes na lista de instâncias de processo do nodo selecionado, representados na consulta pelos parâmetros a₁, a₂, a₃, ..., a_n.

5.2.2. Caso de Uso Alternar Análise KPI → MUW

Este caso de uso representa a mudança da análise quantitativa para a estrutural. A interface do protótipo que realiza este caso de uso é mostrada na Figura 20. Após visualizar e analisar um comportamento de desempenho mediante a aplicação de um KPI, o usuário deseja investigar o fluxo de execução das instâncias de atividades que ocasionaram tal comportamento. Para isso, o usuário seleciona na interface a coluna do gráfico de barras que representa seu interesse de análise. No exemplo ilustrado pela Figura 20, o usuário sumariza todas as instâncias de processos utilizando a métrica *Count of Process Instances by Initiator Resource*. Nessa métrica as instâncias de processos são agrupadas pelo recurso (do tipo humano ou computacional) que a iniciou, e.g. *Resource A*, *Resource B* e *Resource C*. O usuário seleciona a barra correspondente ao seu interesse, exemplificado pela barra correspondente ao recurso B, conforme ilustrado pela Figura 20.

Após a seleção, é gerado um *log* de execução envolvendo todas as atividades executadas pelas instâncias de processo que foram iniciadas pelo Recurso B. Para a geração do *log* de execução é realizada a consulta SQL ilustrada pela Figura 21.

A partir do *log* gerado é construída uma nova árvore agregada (Seção 2.3.1) contendo os fluxos de execução de processo de negócio envolvendo apenas as instâncias de processos iniciadas pelo recurso B. A Figura 22 ilustra a visualização e análise da árvore agregada gerada. Observa-se que todas as 524 instâncias (root – I:524) executaram a seqüência de atividades 1-2. Destas, 327 seguiram a seqüência 3-4-5-.... O restante dos outros fluxos são visualizados com o uso da barra de rolagem.

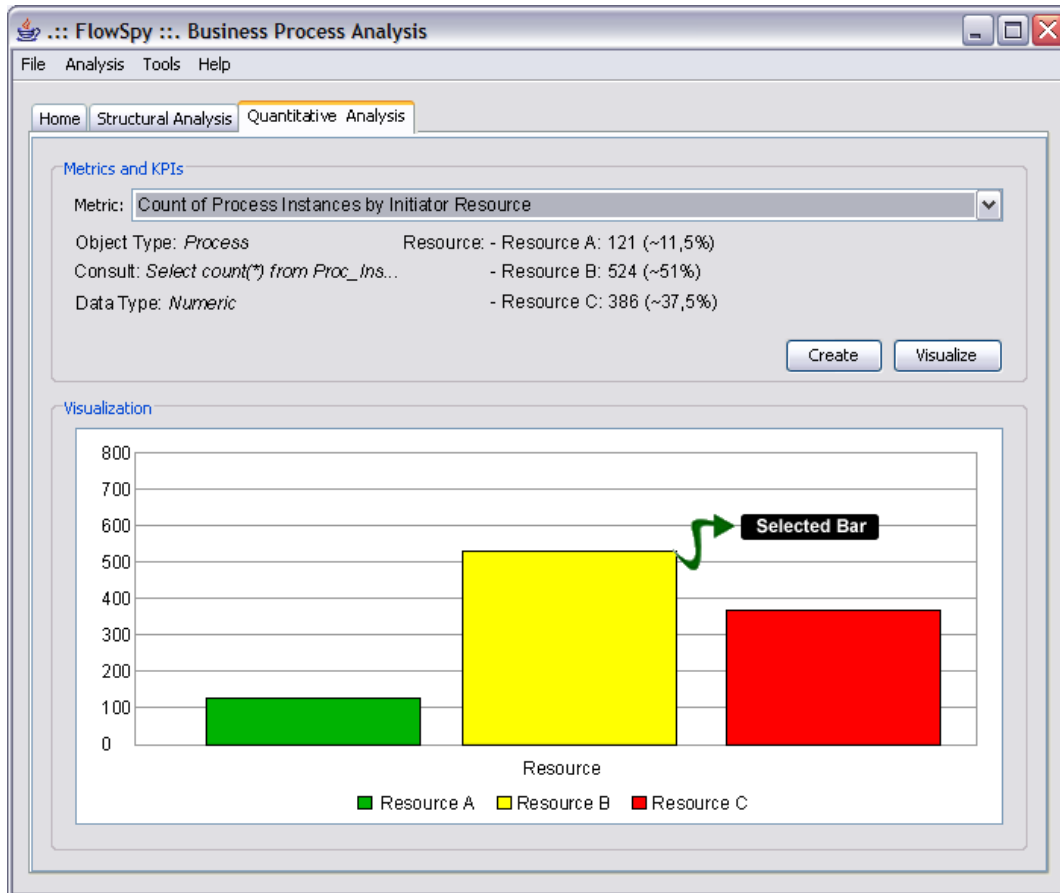


Figura 20 – Análise quantitativa KPI→MUW.

SQL: Geração do Log de Execução

1. Select Fact_Instance_Activity.id_Process, Fact_Instance_Activity.id, Fact_Instance_Activity.time
2. From Fact_Instance_Activity, Fact_Instance_Process
3. Where Fact_Instance_Activity.id_Process = Fact_Instance_Process.id and Fact_Instance_Activity.id = Fact_Instance_Process.Root_Activity and Fact_Instance_Activity.Resource = 'Resource B'
4. Group by Fact_Instance_Activity.id_Process, Fact_Instance_Activity.id, Fact_Instance_Activity.time
5. Order by Fact_Instance_Activity.time;

Figura 21 – SQL para análise quantitativa.

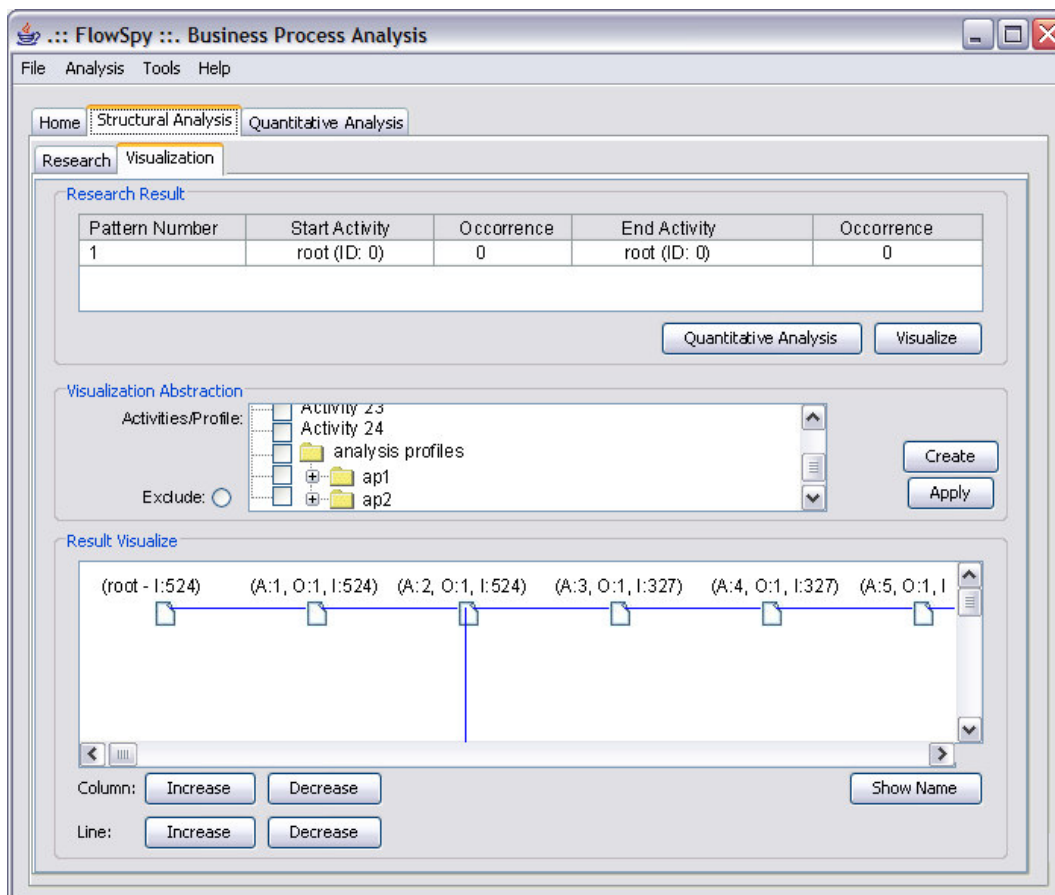


Figura 22 – Análise estrutural KPI→MUW.

5.2.3. Caso de Uso Definir Perfil de Análise

O usuário define um conjunto de atividades que representam o seu interesse de análise. O perfil definido pode ser usado posteriormente pelo gestor de abstração. Nesse caso de uso, o usuário visualiza todas as atividades existentes no *log* de execução e de forma iterativa vai selecionando as atividades que devem compor o perfil. A Figura 23 ilustra a dois estados da interface para criação de perfis de análise.

Para criar um perfil de análise o usuário deve selecionar uma atividade do conjunto existente no campo *Select Nodes* (Figura 23, indicador 1). Inicialmente são disponibilizadas todas as atividades presentes nos fluxos de execução. Após selecionar um nodo (atividade 13 nesse exemplo), o usuário deve inseri-la no seu perfil clicando no botão *Insert* (Figura 23, indicador 2), ao fazer isso a atividade 13 é incluída no campo *Selected Nodes* (Figura 23, indicador 3), e o campo *Select Nodes* (Figura 23, indicador 4) é atualizado

para conter apenas as atividades executadas após qualquer atividade do conjunto *Selected Nodes* (i.e. os nodos que podem ser imediatamente executados a partir dele).

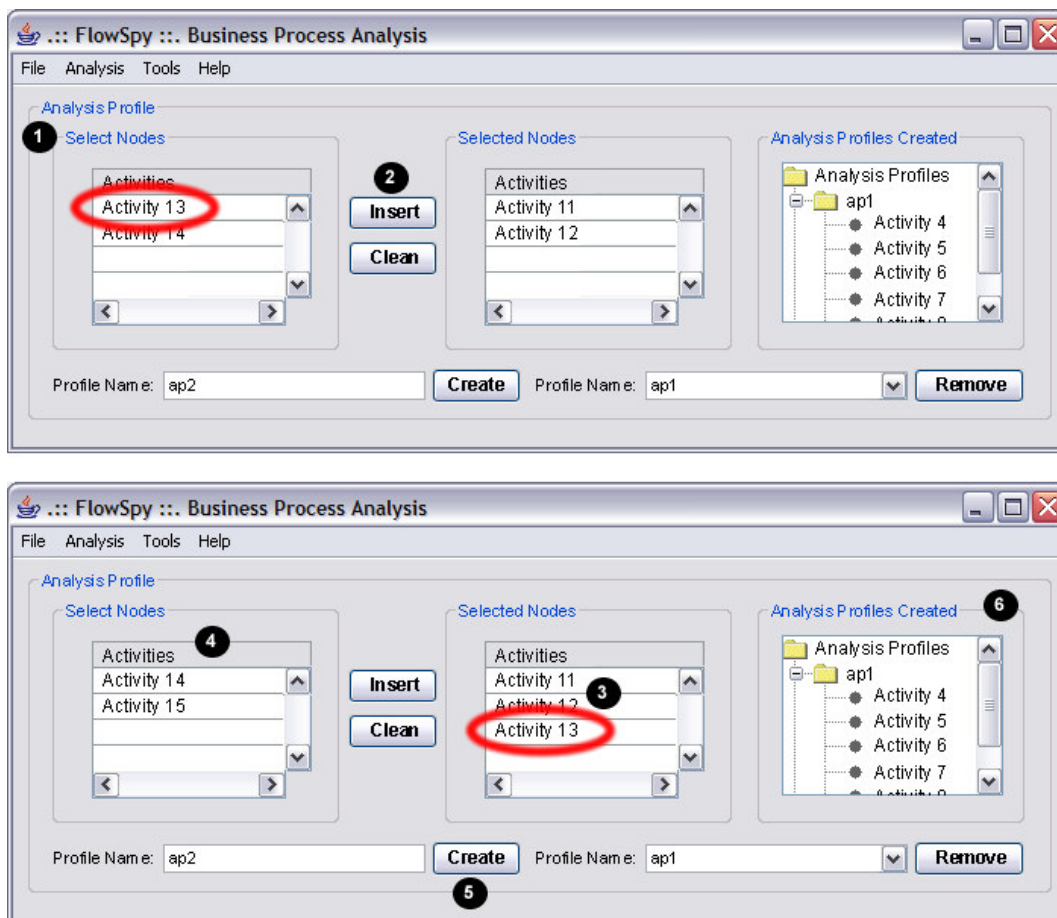


Figura 23 – Caso de uso definir perfil de análise.

A verificação das atividades subseqüentes é feita com base no *log* de execução ou modelo do processo. Para esse exemplo, toma-se como base o modelo de processo apresentado na Figura 10 da Seção 4.1.4.

O usuário repete esses passos até montar o perfil de análise desejado. Por fim, o usuário cria o perfil de análise contendo as atividades do campo *Selected Nodes* ao pressionar o botão *Create* (Figura 23.5). Ele pode verificar este ou outros perfis criados no campo *Analysis Profiles Created* (Figura 23.6).

5.2.4. Caso de Uso Filtrar Log

Este caso de uso representa uma das formas de abstração disponíveis pelo gestor de abstração. Sua função é reduzir o domínio de análise pela exclusão de atividades ou junção

de um conjunto de atividades em um bloco único de execução. Como resultado, tem-se a geração de uma árvore agregada menor contendo fluxos com apenas atividades de interesse.

Para realizar a filtragem do *log*, o usuário deve selecionar as atividades que não sejam de interesse, para que não façam parte da árvore agregada a ser construída. A Figura 24 ilustra este caso de uso.

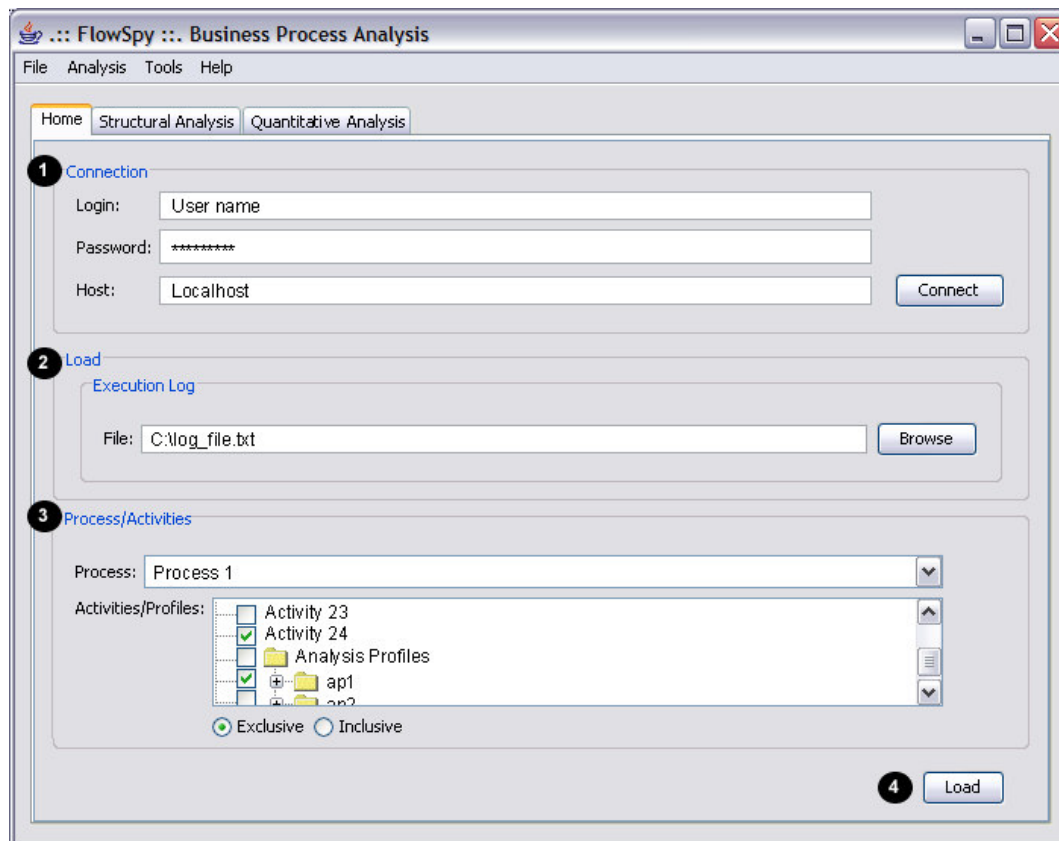


Figura 24 – Caso de uso filtrar log de execução.

Primeiramente o usuário deve conectar-se ao banco de dados informando seu *user name*, *password* e *host* (Figura 24.1) para, a partir do modelo analítico, gerar o *log* de execução e obter informações referentes as atividades existentes. Outra forma de obter tais informações é localizando diretamente um *log* contendo o fluxo de execução das instâncias de processos (Figura 24.2). Após, o usuário seleciona as atividades ou perfis de análise, quando disponível, e define o tipo de filtragem: exclusiva ou inclusiva. A Figura 24.3 ilustra a seleção de uma filtragem do tipo exclusiva da atividade 24 e perfil de análise ap1 (definido na Seção 4.1.4 como PA1). Para finalizar, o usuário pressiona o botão *load* (Figura 24.4) e a árvore agregada é gerada sem a presença da atividade e perfil de análise selecionado. A única diferença da filtragem do tipo inclusiva é que o usuário deve selecionar a opção *inclusive* ao

invés de *exclusive*. Neste caso, o *log* seria gerado contendo apenas as instâncias de atividades selecionadas, i.e. atividade 24 e atividades do perfil de análise ap1.

5.2.5. Caso de Uso Abstrair Visualização

Este caso de uso representa a necessidade de reduzir ou aumentar a quantidade de atividades/nodos presentes no fluxo, pela generalização ou especialização dos padrões minerados, respectivamente. Para isso, o usuário, ao analisar um padrão gerado pela técnica de mineração seqüencial do uso da *web* (árvore de execução), deve selecionar o conjunto de atividades ou perfis de análise que não seja de interesse analítico, e assim, excluir ou agregar esse conjunto do padrão analisado clicando no botão *Apply*.

A Figura 25 ilustra esse caso de uso. O usuário visualiza e analisa o padrão representado pela Figura 17, que mostra a seqüência de atividades executadas para ocasionar a notificação de não execução (atividade de identificador 20) do processo representado pela Figura 10, e não consegue chegar a uma conclusão, pois o padrão possui fluxos de execução envolvendo inúmeras atividades. Desta forma, o usuário seleciona, no campo *Visualization Abstraction*, as atividades ou perfis de análise que deseja excluir ou agregar. No exemplo da Figura 25, o usuário seleciona o perfil de análise ap1. Consequentemente, os fluxos que possuem atividades de ap1 têm suas atividades agregadas em um único nodo de execução (A: ap1). Por *default*, os nodos são agregados.

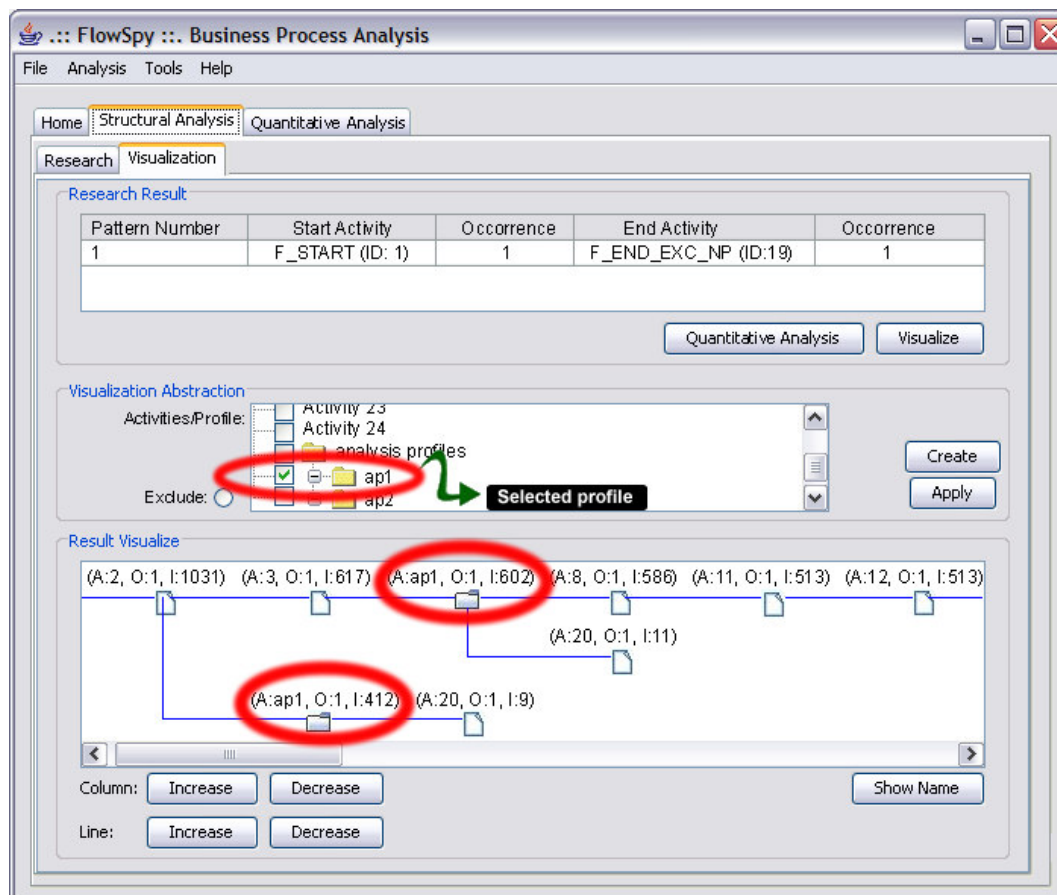


Figura 25 – Caso de uso abstrair visualização do tipo agregação.

Para realizar a abstração de visualização do tipo exclusão, o usuário deve selecionar a opção *Exclude* do campo *Visualization Abstraction*. Ao invés das atividades serem agregadas em um bloco único de execução, elas são simplesmente retiradas do fluxo. A Figura 26 ilustra essa situação. O usuário seleciona o perfil de análise ap1 e a opção *Exclude*. Ao pressionar o botão *Apply*, o usuário faz com que as atividades de ap1 sejam excluídas do fluxo.

Já para voltar ao padrão original, o usuário deve desmarcar as atividades e perfis de análise selecionados e pressionar o botão *Apply* novamente. Nesse exemplo, o padrão voltaria a ser o representado pela Figura 17.

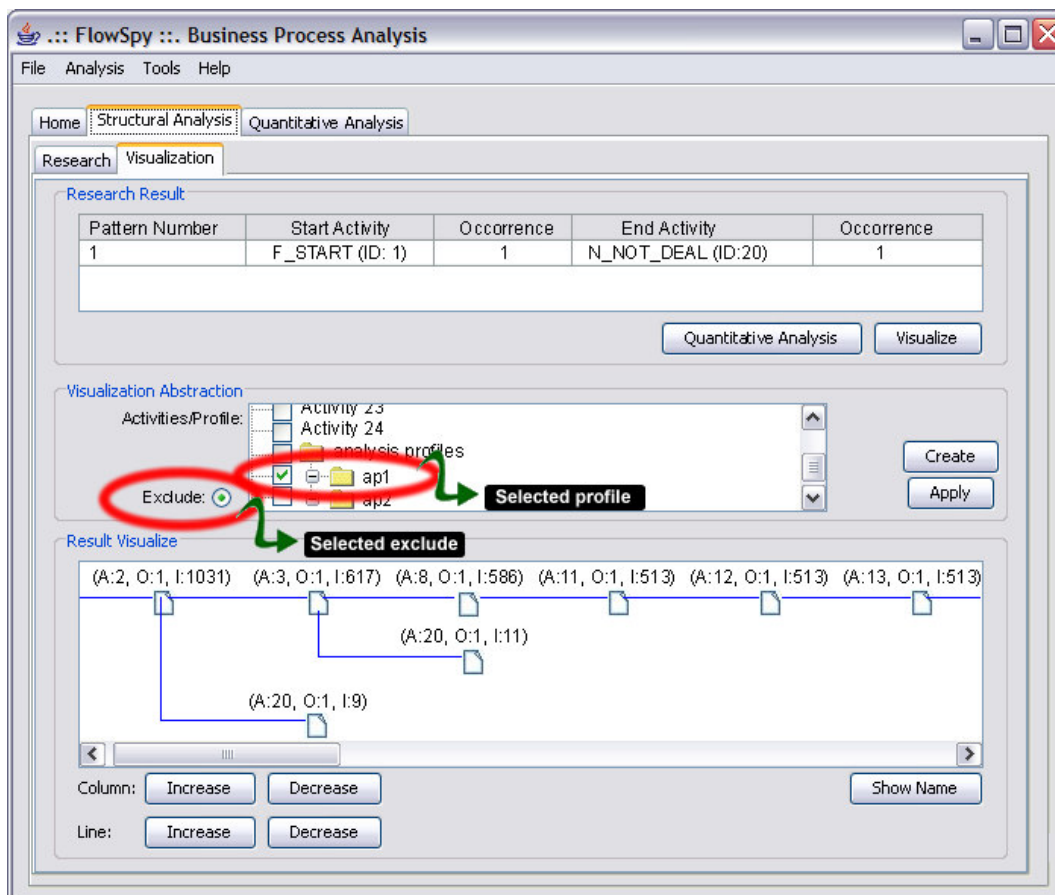


Figura 26 – Caso de uso abstrair visualização do tipo exclusão

5.3. Considerações

Este capítulo apresentou, na forma de casos de uso, experimentações realizadas para analisar os componentes do ambiente proposto neste trabalho. Dados os resultados, verificou-se que o ambiente proposto de fato possibilita integrar diferentes técnicas de análise de processos de negócio. Com efeito, este recurso de integração de dados ou complementação de informação facilita a análise de comportamentos de PN de interesse do usuário. Essa sinergia faz com que o domínio de análise possa ser diminuído e o resultado seja cada vez mais específico a um número reduzido de instâncias de processos.

Os recursos de abstração introduzidos para facilitar a análise e visualização de processos complexos demonstraram ser realmente essenciais, por reduzir ou excluir, do domínio de análise, o conjunto de atividades desnecessárias ao interesse do usuário, facilitando seu trabalho e qualificando suas decisões.

6. CONCLUSÕES E TRABALHOS FUTUROS

Realizando uma análise mais profunda no cenário atual das organizações, conforme apresentado ao longo deste trabalho, fica clara a necessidade de ambientes de BPM que não estejam focados apenas na gerência e automação de processos de negócio. São de fundamental importância a capacidade de análise sobre dados históricos, monitoração (pró-atividade) sobre dados de execução, acesso à informação sob diferentes óticas do negócio, bem como a integração de dados (processo x negócio) dispersos entre os diferentes sistemas de TI presentes na organização.

Os principais objetivos de BPA, nova vertente de BPM, são a obtenção de um maior controle das instâncias de processos e redução do tempo entre a ocorrência de um evento e a adequada tomada de decisão pela pessoa responsável, possibilitando uma maior mensuração e percepção do comportamento da execução dos processos, atividades e recursos, e conseqüentemente a redução de custos, aumento da qualidade do produto e satisfação dos clientes.

O estudo dos trabalhos relacionados realizado no Capítulo 3 permitiu ter uma visão do que é e não é abordado pelas propostas existentes (Quadro 2). Uma característica presente em todas as propostas pesquisadas é a possibilidade de se obter informações de medidas quantitativas de desempenho de acordo com métricas e KPIs definidas para o alcance dos objetivos da empresa. Van der Aalst em [AAL05] apresenta uma abordagem com o uso da técnica de mineração de seqüências, a qual possibilita a comparação e alinhamento do processo de negócio. Contudo, esses tipos de análises não dão suporte à análise e visualização do fluxo das atividades de um processo. A análise e a monitoração do fluxo são importantes para se ter um maior controle da execução das instâncias de processos, possibilitando encontrar e prever comportamentos anormais de atividades (fluxos indesejáveis), e.g. execução de atividades não previstas ou execução de caminhos alternativos. Também, observa-se a falta de mecanismos para integrar os dados de diferentes técnicas de análise e monitoração, não havendo um complemento de informação entre as diferentes técnicas, o que é essencial para compreender determinados comportamentos ou encontrar as causas de anomalias.

O trabalho apresentado por Tristão et al. em [TRI06], o qual utiliza a ferramenta *WebPath* de mineração e visualização de caminhos da *web* no contexto de PN, serviu para verificar a contribuição da mineração seqüencial para análise do comportamento dos

processos de negócio. Com esse tipo de análise, além de acompanhar de forma precisa a execução das atividades, pôde-se identificar o nível de convergência (probabilidade de execução) entre as atividades, verificar instâncias em execução e encontrar possíveis anomalias de execução do processo.

A proposta do ambiente de BPA destinado a integração de dados e informação provenientes da técnica de mineração sequencial (*WebPath*) e análise quantitativa de processos (BPI) demonstrou, por meio de experimentações envolvendo os componentes desenvolvidos e ilustrados através dos casos de uso, que viabilizar a sinergia entre as duas técnicas proporciona uma redução do domínio de análise e aumenta o poder informacional, reforçadas pelo acréscimo de recursos de abstração propostos para a mineração de seqüências. Esse ambiente apóia a representação e manipulação gráfica dos fluxos das instâncias de processos de negócio, possibilita a aplicação de recursos analíticos e métricas aos processos, atividades e/ou recursos para identificar anomalias de execução e de desempenho, bem como obter diferentes visões do processo de acordo com o interesse de análise.

Além desta tese de dissertação, foram publicados mais três trabalhos científicos: um artigo no V Workshop de Teses e Dissertações em Bancos de Dados (WTDBD 2006); um artigo na II Escola Regional de Banco de Dados (ERBD 2006) [TRI06]; e um resumo do projeto de pesquisa na I Mostra de Pesquisa da Pós-Graduação da PUCRS.

Essa proposta permite que futuramente sejam incorporadas novas técnicas para análise, monitoração e visualização de processos de negócio, bem como propor e desenvolver os demais componentes do ambiente, e.g. modelos de dados com suporte a dados do processo e do negócio e técnicas alternativas de ETL para a redução da latência de dados. A implementação desse ambiente também pode ser realizada futuramente.

Outra possibilidade de trabalho futuro identificada nesta proposta é com relação a modelos de predição. Conforme Grigori et al. [GRI04], a geração de modelos de predição com o uso da técnica de classificação não é um processo trivial. Nessa técnica deve ser gerada uma árvore de decisão para cada estado de execução de um processo. Por outro lado, a estrutura de dados utilizada para representar os diferentes fluxos (árvore agregada) já representa todos os fluxos de execução existentes no processo. Dessa forma, necessita-se apenas evoluir a estrutura de modo a suportar as informações quantitativas para poder comparar com o desempenho de cada instância de processo em execução.

REFERÊNCIAS BIBLIOGRÁFICAS

- [AAL05] AALST, W.M.P. van der. Business Alignment: Using Process Mining as a Tool for Delta Analysis and Conformance Testing. **Requirements Engineering Journal**, London, v. 10, n. 3, p. 198-211, Nov. 2005.
- [AAL03] _____; HOFSTEDE, A. H.M. ter; WESKE, M. Business Process Management: A survey. **Lecture Notes in Computer Science**, Berlin, v. 2678, p. 1-12, Jan. 2003.
- [CAS05a] CASATI, F. Industry Trends in Business Process Management: Getting Ready for Prime Time. In: INTERNATIONAL WORKSHOP ON DATABASE AND EXPERT SYSTEMS APPLICATIONS (DEXA'05), 16., 22-26 Aug. 2005, Copenhagen. **Proceedings...** Copenhagen: IEEE Computer Society, Aug. 2005, p. 903-907.
- [CAS05b] CASTELLANOS, M.; CASATI, F.; MING-CHIEN S.; Dayal, U. iBOM: A Platform for Intelligent Business Operation Management. In: INTERNATIONAL CONFERENCE ON DATA ENGINEERING (ICDE 2005), 21., 5-8 Apr. 2005, Tokyo. **Proceedings...** Tokyo: IEEE Computer Society, Apr. 2005, p. 1084-1095.
- [CAS02] CASATI, F. **Intelligent Process Data Warehouse for HPPM 5.0**: HPL-2002-120, HP Laboratories Palo Alto: Software Technology Laboratory, 26 Apr. 2002.
- [CHA02] CHANG, S.; JAECKEL, C. **Oracle Workflow Guide**: Release 2.6.2, San Francisco: Oracle Corp., Mar. 2002, v. 1.
- [COO99] COOLEY R., MOBASHER B. AND SRIVASTAVA J. Data preparation for mining world wide Web browsing patterns. **Knowledge and Information Systems**, London, v. 1, n 1, p. 5-32, Feb. 1999.
- [DRE03] DRESNER, H. Business Activity Monitoring: BAM Architecture. In: GARTNER SYMPOSIUM ITXPO, Nov. 2003, Cannes. **Proceedings...**Cannes: Gartner Inc., Nov. 2003, p. 1-20.
- [FAY96] FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. The KDD Process for Extracting Useful Knowledge from Volumes of Data. **Communications of the ACM**, New York, v.39, n. 11, p. 27-34, Nov. 1996.

- [GOL04] GOLFARELLI, M., RIZZI, S.; CELLA, L. Beyond data warehousing: what's next in business intelligence? In: ACM INTERNATIONAL WORKSHOP ON DATA WAREHOUSING AND OLAP, 7., Nov. 2004, Washington. **Proceedings...** New York: ACM Press, Nov. 2004, p. 1-6.
- [GAR05] GARCIA, R. S.; RUIZ, D. D. Pré-Processamento de Dados para Descoberta de Conhecimento em Processos de Workflow Modelados sobre Plataforma Oracle. In: ESCOLA REGIONAL DE BANCO DE DADOS (ERBD 2005), 1., abr. 2005, Porto Alegre. **Anais...** Porto Alegre: Instituto de Informática da UFRGS, abr. 2005, p. 25-30.
- [GRI04] GRIGORI, D., CASATI, F., CASTELLANOS, M., DAYAL, U., SAYAL, M.; SHAN, M. C. Business Process Intelligence. **Computers in Industry**, Elsevier, v. 53, n. 3, p. 321-343, Apr. 2004.
- [HAN01] HAN, J.; KAMBER, M. **Data Mining: Concepts and Techniques**. San Francisco: Morgan Kaufmann, 2001. 550 p.
- [HEL02] HELLINGER, M. AND FINGERHUT, S. Business Activity Monitoring: EAI Meets Data Warehousing. **EAI Journal**, Dallas, p. 18-21, July 2002.
- [LIS04] LIST, B.; MACHACZEK, K. Towards a Corporate Performance Measurement System. In: ACM SYMPOSIUM ON APPLIED COMPUTING, 19., 14-17 Mar. 2004, Nicosia. **Proceedings...** New York: ACM Press, Mar. 2004, p. 1344-1350.
- [MEL04a] MELCHERT, F.; WINTER, R. The Enabling Role of Information Technology for Business Performance Management. In: IFIP INTERNATIONAL CONFERENCE ON DECISION SUPPORT SYSTEMS (DSS 2004), 1-3 July 2004, Prato. **Proceedings...** Prato: Monash University IT, July 2004, p. 535-546.
- [MEL04b] _____; KLESSE, M. Aligning Process Automation and Business Intelligence to Support Corporate Performance Management. In: AMERICAS CONFERENCE ON INFORMATION SYSTEMS, 10., 5-8 August 2004, New York. **Proceedings...** New York: Stevens Institute of Technology, 5-8 August 2004, p. 4053-4063.
- [OMG04] **UML 2.0 SuperStructure Specification**: ptc/04-10-02, Needhan: Object Management Group, Oct. 2004, p. 804.
- [OPE06] **Java Open Source**: workflow engine based on XPDL. Desenvolvido por Enhydra Shark project. Disponível em:
<<http://www.enhydra.org/workflow/shark/>>. Acesso em: 15 nov. 2006.

- [SAY01] SAYAL, M.; CASATI, F.; DAYAL, U.; SHAN, M. C. **Business Process Cockpit**: HPL-2001-313, HP Laboratories Palo Alto: Software Technology Laboratory, 28 Nov. 2001.
- [SCH04] SCHIEFER, J.; JENG, J.; KAPOOR, S.; CHOWDHARY, P. Process information factory: a data management approach for enhancing business process intelligence. In: IEEE INTERNATIONAL CONFERENCE ON E-COMMERCE TECHNOLOGY (CEC'04). 6-9 July 2004, San Diego. **Proceedings...** San Diego: IEEE Computer Society, July 2004, p. 162-169.
- [SPI00] SPILIOPOULOU, M. Web Usage Mining for Site Evaluation: Making a site better fit its users, **Communications of the ACM**, New York, v. 43, n. 8, p. 127-134, Aug. 2000.
- [SPI98] _____; FAULSTICH, L. C. WUM: a Web Utilization Miner. In: INTERNATIONAL WORKSHOP ON THE WEB AND DATABASES, 1., Mar. 1998, Valencia. **Proceedings...** Valencia: Universitat Politecnica de Valencia, Mar. 1998, p. 109-115.
- [SRI00] SRIVASTAVA, J.; COOLEY, R.; DESHPANDE, M.; TAN, P. N. Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data. **Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD Explorations)**, v. 1, n. 2, p. 12-23, Jan. 2000.
- [TRI06] TRISTÃO, C.; BECKER, K. Aplicação de Técnicas de Mineração do Uso da Web para Análise de Processos de Negócio: um Estudo de Caso. In: ESCOLA REGIONAL DE BANCO DE DADOS (ERBD 2006), 2., 6-8 abr. 2006, Passo Fundo. **Anais...** Passo Fundo: UPF, abr. 2006, p. 57-62.
- [TRI05a] TRISTÃO, C. **Um Estudo sobre Inteligência de Processos de Negócio**. nov. 2005, 70 f. Trabalho Individual II (Mestrado em Ciência da Computação)-Faculdade de Informática, PUCRS, Porto Alegre, 2005.
- [TRI05b] _____. **Um Estudo Sobre Monitoramento e Análise de Processos de Negócio**. jun. 2005, 62 f. Trabalho Individual I (Mestrado em Ciência da Computação)-Faculdade de Informática, PUCRS, Porto Alegre, 2005.
- [TRI04] _____.; TEIXEIRA, V. P.; BECKER, K. WebPath: Uma Ferramenta para Mineração e Visualização de padrões de Navegação do Uso da Web. In: SIMPOSIO BRASILEIRO DE SISTEMAS DE INFORMAÇÃO (SBSI 2004), 1., 13-14 out. 2004, Porto Alegre. **Anais...** Porto Alegre: PUCRS, out. 2004, p. 103-110.

[WES04] WESKE, M.; VAN DER AALST, W.; VERBEEK, H. Advances in Business Process Management. **Data & Knowledge Engineering**, Elsevier, v. 50, n. 1, p. 1-8, Jan. 2004.

[WMC99] **The Workflow Management Coalition Specification - Terminology & Glossary**: WFMC-TC-1011, Winchester: Workflow Management Coalition, Feb. 1999, p. 65.