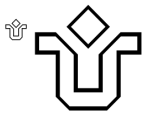


UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

Story Mining: Elicitação de Processos de Negócio a partir de Group Storytelling e Técnicas de Mineração de Texto

João Carlos de Almeida Rodrigues Gonçalves

Orientadora
Flávia Maria Santoro

Co-Orientadora
Fernanda Araujo Baião Amorim

RIO DE JANEIRO, RJ – BRASIL
Abril de 2010

STORY MINING: ELICITAÇÃO DE PROCESSOS DE NEGÓCIO A PARTIR DE GROUP STORYTELLING E TÉCNICAS DE MINERAÇÃO DE TEXTO

João Carlos de Almeida Rodrigues Gonçalves

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

Aprovada por:

Profa. Flávia Maria Santoro, D.Sc. - UNIRIO

Profa. Fernanda Araujo Baião Amorim, D.Sc. - UNIRIO

Profa. Renata Mendes de Araujo, D.Sc. - UNIRIO

Prof. Alexandre Gonçalves Evsukoff, Dr. – COPPE/UFRJ

RIO DE JANEIRO, RJ – BRASIL

Abril de 2010

Gonçalves, João Carlos de Almeida Rodrigues.
G635 Story mining : elicitação de processos de negócio a partir de group storytelling e técnicas de mineração de texto / João Carlos de Almeida Rodrigues Gonçalves, 2010.
175f.

Orientador: Flávia Maria Santoro.
Coorientador: Fernanda Araujo Baião Amorim.
Dissertação (Mestrado em Informática) – Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro, 2010.

1. Modelo de negócio. 2. Tecnologia da informação. 3. Sistema de informação. 4. Colaboração. 5. Gestão do conhecimento. 6. Mineração de texto. I. Santoro, Flávia Maria. II. Amorim, Fernanda Araujo Baião. III. Universidade Federal do Estado do Rio de Janeiro (2003-). Centro de Ciências Exatas e Tecnologia. Curso de Mestrado em Informática. IV. Título.

CDD – 005.5

A meus pais e amigos.

Agradecimentos

Agradeço a todos que tornaram possível minha caminhada profissional chegar a este ponto.

Um agradecimento muito especial aos meus pais, que me apoiaram desde o início desta trajetória e aos meus amigos, pela compreensão ante a minha falta de atenção por causa do trabalho árduo do mestrado.

Em especial agradeço às minhas orientadoras, Flávia Maria Santoro e Fernanda Araújo Baião Amorim pela oportunidade de trabalhar com pessoas tão competentes e pelo seu apoio na hora certa, a paciência com minhas dúvidas e pelo aprendizado intenso que foram estes anos de trabalho de pesquisa do Mestrado. Sua competência e rigor, suas sugestões acertadas e o estímulo constante a enfrentar os desafios da pesquisa foram fundamentais para este trabalho.

Agradeço a todo o Corpo Docente do PPGI, que me proporcionou uma experiência única de aprendizado e de troca de idéias. Em especial, além de minhas orientadoras, aos professores Márcio Barros, Renata Araujo e Sean Siqueira, pela oportunidade de aprender em suas disciplinas a base de áreas de conhecimento que foram vitais para esta dissertação.

Agradeço também ao Prof. Alexandre Gonçalves Evsukoff, pelas valiosas sugestões relativas a questões da área de Mineração de Texto.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa de fomento concedida.

Agradeço também aos colegas do PPGI, que proporcionaram um ambiente agradável durante todo o mestrado.

À colega de mestrado Lúcia Castro pelos *insights* sobre a estrutura da língua portuguesa.

A Raphael Perret, pela ajuda com a ferramenta TellStory original.

À Cláudia Cappelli, pelas sugestões valiosas para o método Story Mining.

Uma lembrança especial a todos os companheiros do laboratório NP2TEC, em especial a Alexandre Souza, Bruna Diirr, Clarissa Romeiro, Douglas Machado, Daniel Serrano, Daniel Bizzi, Diego Duarte, Marcos Veloso, Raphael Tiago, Sérgio Puntar, Licia Nascimento, dentre outros que criaram um ambiente descontraído e agradável para a pesquisa.

Ao grupo de pesquisa NP2TEC, minha sincera gratidão pela oportunidade de integrar um grupo seleto de profissionais de alta qualidade.

Aos participantes de meus estudos de casos, agradeço pela colaboração. Sem as pessoas envolvidas, pouco poderia ter sido feito nesta pesquisa.

Um agradecimento aos técnicos administrativos Ercília Moreira e Alessandra Nascimento pelo trabalho sério e conversas descontraídas.

Por fim, agradeço a todos que possa ter esquecido de citar aqui.

Gonçalves, João Carlos de Almeida Rodrigues. Story Mining: Elicitação de Processos de Negócio a partir de Group Storytelling e Técnicas de Mineração de Texto**.** UNIRIO, 2010. 175 páginas. Dissertação de Mestrado. Departamento de Informática Aplicada, UNIRIO.

RESUMO

A Modelagem de Processos de Negócio tem sido adotada por organizações em variados contextos, como análise de melhores práticas, design de sistemas e arquiteturas de Tecnologia da Informação.

Uma das fases mais importantes desta tarefa é a Elicitação de Processos de Negócio, que busca adquirir o conhecimento envolvido no processo e tornar possível sua utilização pelo analista. Embora influencie diretamente a qualidade do modelo final, o processo de elicitação ainda é bastante complexo e passível de erros que podem comprometer a qualidade do modelo final.

Dois pontos principais podem ser apontados, relacionados diretamente com estas dificuldades: A expressão de conhecimento por parte dos envolvidos com o processo e a interpretação das informações adquiridas pelo analista. Diversos problemas como a influência do viés do analista na interpretação e informações incompletas por parte dos stakeholders podem também ser percebidos nas práticas tradicionais da literatura.

O estudo apresenta uma nova abordagem que visa solucionar os problemas citados, explorando para tal fim o uso de narrativas colaborativas (Group Storytelling) associadas com técnicas de Mineração de Texto para suporte à Elicitação. O novo método, denominado StoryMining, teve sua viabilidade avaliada por estudos de casos em organizações e processos de diferentes perfis, além da análise de modelos de processos já existentes.

Palavras-chave: Modelagem de Processos de Negócio, Elicitação de Processos de Negócio, Gestão de Conhecimento, Mineração de Texto.

ABSTRACT

Business process modeling has been adopted by organizations in various contexts, such as analysis for best practices, system design and Information Technology architecture.

One of the most important phases of the modeling task is the process elicitation, aiming at acquiring process knowledge and making it available for the analyst. Besides being a major influence for the quality of the final model, the elicitation process is still costly and complex.

Two main issues can be pointed, which are directly related with these difficulties: Knowledge expression by the people involved with the process and interpretation of the information acquired by the analyst. Several problems, such as the influence of the analyst bias and incomplete information supplied by the stakeholders, may occur when applying the traditional practices present at the literature.

This study presents a novel approach, which aims at solving the described problems through the use of collaborative narratives (Group Storytelling) together with Text Mining techniques for elicitation support. The new method, Story Mining, was applied in case studies of organizations using different profiles and processes, also considering the analysis of pre-existent process models.

Keywords: Business Process Modeling, Business Process Elicitation, Knowledge Management, Text Mining

Sumário

Capítulo 1 – Introdução 15

1.1 Motivação 15

1.2 Caracterização do Problema 16

1.3 Metodologia de Pesquisa 17

1.4 Hipótese de Pesquisa 18

1.5 Objetivos da Dissertação 18

1.6 Organização da Dissertação 19

Capítulo 2 – Elicitação e Modelagem de Processos de Negócio 20

2.1. Definição 20

2.2. O conhecimento na elicitação de processos de negócio 20

2.3. A Abordagem Típica de Elicitação de Processos de Negócio 23

2.4. Extração automática de processos de negócio (*Process Mining*) 24

2.5. Trabalhos relacionados 25

Capítulo 3 – Group Storytelling 30

3.1. Introdução 30

3.2. As Histórias 30

3.3. A técnica Group Storytelling 31

3.4. As histórias e a memória organizacional 32

3.5. O Processo do Group Storytelling 33

3.6. Fases do Group Storytelling 34

3.7. Elementos da Narrativa 36

3.7.1. Eventos 36

3.7.2. Personagens 37

3.7.3. Documentos 37

3.7.4. Comentários 37

3.7.5. Votações 38
3.7.6. Versões 38
3.8. Group Storytelling aplicada a Processos de Negócio 38
3.9. Dificuldades da manipulação do conhecimento adquirido 40
Capítulo 4 – Mineração de Textos 42
4.1. Definição 42
4.2. Mineração de Textos e Áreas relacionadas 43
4.3. Conceitos Básicos de Mineração de Texto 44
4.4.3.1. Documentos 44
4.4.3.2. Modelos Representativos dos Documentos 45
4.3.2.1. Caracteres 47
4.3.2.2. Palavras 47
4.3.2.3. Termos 47
4.3.2.4. Conceitos 48
4.4.3.3. Coleções de Documentos 49
4.4.3.4. Domínios e Conhecimento Externo 50
4.4. Processo de mineração de textos 50
4.4.1. Pré-Processamento 51
4.4.1.1. Filtragem por elementos básicos e *stopwords* 52
4.4.1.2. Ocorrência de Palavras 52
4.4.1.3. Resolução de Anáforas e Co-referência 53
4.4.1.4. POS-Tagging (Part-Of-Speech Tagging) 53
4.4.1.5. Parsing Sintático 54
4.4.1.6. Shallow Parsing 55
4.4.2. Operações de Mineração 56
4.4.2.1. Reconhecimento de Entidades Nomeadas 58
4.4.2.2. Extração de Elemento de Template 59
4.4.2.3. Extração de Relacionamento de Template 60
4.4.2.4. Extração de Co-Referência 60
4.4.2.5. Extração de Cenário de Template 61
4.4.3. Pós-Processamento 62
4.4.3.1. Apresentação de Resultados 62
4.5. Considerações sobre Mineração de Texto aplicada aos Processos de Negócio 63
Capítulo 5 – O Método Story Mining 65
5.1. Requisitos de um método para elicitação e modelagem de processos 65
5.2. Papéis envolvidos 67
5.3. Fases do método 68

5.4.1. Primeira Fase: Construção coletiva de histórias 68
5.4.2. Segunda Fase: Mineração de histórias 69
5.4.2.1. Tokenização 71
5.4.2.2. Análise Léxica e Morfológica 71
5.4.2.3. Análise Sintática 72
5.4.2.4. Análise de Domínio 72
5.4.3. Terceira Fase: Construção de representação formal 73
Capítulo 6 – Ambiente de Suporte ao Método StoryMining 75
6.1. Introdução 75
6.2. A Ferramenta ProcessTeller 76
6.2.1. Classificação das narrativas utilizando Grupos de Histórias 77
6.2.2. Vinculação de documentos aos eventos da história 78
6.2.3. Detecção de Personagens ao nível de eventos da história 79
6.2.4. Evento inicial da história 80
6.3. O Sistema de Workflows Científicos VisTrails aplicado à MT 80
6.3.1. Descrição do Workflow do Método 82
6.3.1.1. Primeira parte: Processamento dos documentos das histórias 83
6.3.1.2. Segunda parte: Processamento dos eventos das histórias 83
6.3.1.3. Terceira parte: Extração de Informação 84
6.4. Visualização de Proto-Modelo em notação BPMN pelo BizAgi 84
Capítulo 7 – Estudos de Caso e Análise de Resultados 87
7.1. Metodologia 87
7.2. Estudo de Caso 1: Processo de Inscrição em Disciplinas DIA/UNIRIO 89
7.2.1. Descrição 89
7.2.2. Análise 91
7.3. Estudo de Caso 2: Processo “Gerir Levantamento de Processos” 101
7.3.1. Descrição 101
7.3.2. Análise 103
7.4. Discussão 110
Capítulo 8 – Conclusões e Trabalhos Futuros 113
8.1. Considerações Finais e Trabalhos Futuros 114
8.2. Contribuições Realizadas 116
Referências Bibliográficas 118
Anexo A – Etiquetas do Corpus MAC-MORPHO 127

Anexo B – Gramáticas utilizadas para a Implementação..... 128

Apêndice A – Questionário para os Estudos de Caso..... 129

Apêndice B – Resultados dos Estudos de Caso..... 132

Lista de Figuras

Figura 1 – Espiral do Conhecimento de NONAKA & TAKEUCHI (1997) 22
Figura 2 – Atividades da Engenharia de Requisitos, extraída de HICKEY & DAVIS (2008) 26
Figura 3 – Diagrama de Técnicas de Elicitação de Requisitos (TSUMAKI & TAMAI, 2005) 28
Figura 4 – Estrutura da dinâmica de Group Storytelling (PERRET, 2004) 36
Figura 5 – Método para Group Storytelling aplicado à Elicitação de Processos (SANTORO, BORGES & PINO, 2008) 39
Figura 6 – Processo geral de Mineração de Texto 51
Figura 7 – Processo Geral de um Sistema de Extração de Informação 58
Figura 8 – Método Story Mining 67
Figura 9 – Modelo CREWS (ACHOUR, 1998) adaptado 71
Figura 10 – Exemplo de trecho de texto após a fase de análise sintática 72
Figura 11 – Exemplo de arquivo de log do processo de extração 73
Figura 12 – Método StoryMining com as aplicações utilizadas 76
Figura 13 - A ferramenta ProcessTeller 77
Figura 14: Interface principal do TellStory, com os grupos de histórias 78
Figura 15: Lista de documentos vinculados a um evento específico 79
Figura 16: Tela de aviso para um evento que não contém personagens registrados 80
Figura 17 – Workflow utilizado para o método 82
Figura 18 – Visualização do proto-modelo, mostrando atividades e atores candidatos . 85

Lista de Tabelas

Tabela 1 – Resultado comparativo entre os dois modelos (Gramática I) 91
Tabela 2 – Atividades Extraídas por grupo (Gramática I) 92
Tabela 3 – Atividades do Processo validado (Inscrição em Disciplinas) 93
Tabela 4 – Atividades extraídas relacionadas com o modelo validado (Gramática I) ... 94
Tabela 5 – Atividades Novas extraídas pelo Método (Gramática I) 95
Tabela 6 – Atividades extraídas com o uso da Gramática II..... 97
Tabela 7 – Questionário com a contagem de respostas 99
Tabela 8 – Resultado comparativo entre os dois modelos (Gramática I) 103
Tabela 9 – Atividades do Processo validado (Gerência de Levantamento de Processos) 105
Tabela 10 – Atividades extraídas (Gramática I) relacionadas com o modelo validado 106
Tabela 11– Atividades novas extraídas pelo método (Gramática I)..... 106
Tabela 12 – Atividades extraídas com o uso da Gramática II..... 107
Tabela 13 – Questionário com a contagem de respostas 109

Capítulo 1 – Introdução

Neste capítulo será a apresentada a motivação para esta pesquisa, bem como o problema abordado e a hipótese assumida. Além disso, serão apresentados os objetivos do trabalho e a organização das partes deste documento.

1.1 Motivação

A modelagem de processos tem sido adotada por organizações nos contextos mais diversos em diversos países. Embora seja uma prática importante, permanece sendo bastante complexa e dispendiosa, mesmo após anos de pesquisa na área.

Os modelos de processos de negócio são utilizados para diversos fins, como a revisão de melhores práticas, além de servir como base para a definição de estratégia para Tecnologia da Informação (TI) e para o *design* de sistemas de informação (WESTON, CHATHA, AJAEFOBI, 2004). Pesquisas recentes também sugerem a Arquitetura orientada a Serviços (*Service Oriented Architecture* – SOA) como solução promissora na área (WOODLEY & GAGNON, 2005).

A qualidade de um modelo de processo de negócio é bastante importante, e diretamente relacionada com a representação mais fiel possível da realidade. Caso os modelos não possuam um grau de qualidade suficiente, não serão úteis para os propósitos da organização ou, ainda pior, poderão sugerir e fornecer informações errôneas.

A elicitação de processos busca levantar informações sobre o processo a ser modelado e disponibilizar este conhecimento para uso posterior, principalmente para a tarefa de modelagem de processos de negócio. Dentre as diferentes tarefas envolvidas na modelagem, a elicitação permanece sendo uma das mais importantes, podendo comprometer a qualidade de seu produto final, o modelo de processo de negócio, em caso de falhas e erros.

Dentre as características de um processo de elicitação de qualidade, a questão da externalização do conhecimento se destaca como um dos principais fatores para o seu sucesso. A dificuldade de compartilhar o conhecimento sobre o processo, muitas vezes presente apenas na mente das pessoas envolvidas com suas atividades, é um problema de pesquisa relevante (KALPIC & BRENUS, 2006).

A abordagem mais utilizada na área de elicitação de processos de negócio envolve um analista conduzindo entrevistas com pessoas que realizam as tarefas a serem analisadas e a posterior representação do conhecimento adquirido através de modelos. Neste caso, a qualidade dos modelos depende fortemente da habilidade do analista responsável e das pessoas que descreveram o cenário do processo (DEN HENGST & DE VREEDE, 2004), configurando-se assim, como uma limitação da abordagem tradicional.

Alguns estudos parecem indicar um caminho promissor de extração automática dos modelos de processos através de técnicas de mineração sobre os logs de sistemas de informação que suportam a execução do processo. Esta abordagem é denominada Mineração de Processos ("Process Mining"), que envolve conceitos principalmente das áreas de Mineração de Textos (INGVALDSEN, GULLA *et al.*, 2005) (INGVALDSEN, 2006) e Processamento da Linguagem Natural (GHOSE, KOLIADIS *et al.*, 2007) (SINHA, PARADKAR *et al.*, 2008). A interseção destas áreas com a área de Modelagem e Elicitação de Processos de Negócio se faz justamente de modo a amenizar o volume de informações a serem processadas pelo analista e no suporte direto ao trabalho de modelagem e de interpretação do conhecimento explícito pelos envolvidos e sua posterior transformação em um modelo conciso de processo de negócio.

Neste contexto, na próxima sub-seção é destacado o problema a ser abordado pelo trabalho de pesquisa desta dissertação.

1.2 Caracterização do Problema

Diversas questões em aberto se apresentam na Modelagem de Processos de Negócio, dentre os quais podemos destacar duas: o fato de que o conhecimento sobre um processo geralmente não se encontra na mente de uma única pessoa, mas sim disperso em um grupo de indivíduos, envolvido com as tarefas do processo de negócio (DE VREEDE, SOL, DICKSON *et al.*, 1995). Isto requer que a modelagem e elicitação dos processos de negócio leve em conta todo o grupo de pessoas envolvido (INDULSKA, RECKER, ROSEMANN *et al.*, 2009).

A segunda questão é o viés que um analista pode exercer sobre o conhecimento adquirido, principalmente na abordagem tradicional utilizando entrevistas para elicitação, seja por sua orientação do levantamento de informações ou pela seleção das pessoas a serem abordadas (ALVAREZ, 2002).

Sendo assim, o problema a ser tratado nesta dissertação é: como realizar a elicitação de processos limitando o viés introduzido pelas entrevistas tradicionais, de forma a tornar esta atividade mais eficiente e menos dispendiosa.

1.3 Metodologia de Pesquisa

A pesquisa desta dissertação foi iniciada com uma revisão bibliográfica sobre a área de Modelagem de Processos de Negócio, se estendendo às áreas de pesquisas correlatas, como Elicitação de Conhecimento e Elicitação de Requisitos de Software. A partir deste trabalho inicial, se definiram os problemas a serem abordados e possíveis iniciativas de solução.

Uma nova pesquisa bibliográfica foi realizada, com um enfoque mais específico nos problemas da Elicitação e posteriormente Modelagem de Processos de Negócio. A busca por soluções adicionou a estes esforços a pesquisa por publicações das áreas de Mineração de Texto e de Group Storytelling.

Com o conhecimento da literatura, foi possível propor uma nova abordagem, compará-la com outras soluções existentes e implementá-la como prova de conceito do método proposto.

Mais adiante, foram realizados dois estudos de casos para avaliar a proposta, utilizando organizações e processos de perfis diferentes para comparação.

Deste modo, a metodologia foi composta pelas seguintes etapas:

1ª Etapa: Pesquisa Bibliográfica

- Pesquisa na literatura sobre a área de Modelagem e Elicitação de Processos de Negócio, além de suas áreas correlatas, como a Elicitação de Requisitos de Software e Elicitação de Conhecimento.

- Pesquisa na literatura sobre possíveis soluções para os problemas encontrados, abrangendo as áreas de Mineração de Texto e Group Storytelling.

2ª Etapa: Proposta e Implementação

- Proposta do novo método para Modelagem e Elicitação de Processos de Negócio.
- Implementação do método proposto em um ambiente computacional.

3ª Etapa: Avaliação da Proposta por Estudos de Caso

4ª Etapa: Análise de Resultados

1.4 Hipótese de Pesquisa

A hipótese estabelecida nesta dissertação é a seguinte: "A Modelagem de Processos de Negócio pode ser realizada com o uso de narrativas descritas com pouca influência do analista de negócios".

Com este enfoque, é proposta uma nova abordagem, voltada à redução do possível viés do analista, utilizando técnicas de Group Storytelling e Mineração de Texto.

1.5 Objetivos da Dissertação

O objetivo da pesquisa foi analisar os problemas da elicitação e modelagem de processos de negócio, e explorar uma nova solução para a área, visando aprimorar a aquisição de conhecimento sobre o processo a ser modelado e facilitar a construção de um modelo de processo de maior qualidade.

Para este fim, os seguintes objetivos foram perseguidos durante o trabalho de pesquisa:

- Proposta de um novo método para a área, que torne possível mitigar os problemas encontrados e reforçar as qualidades desejadas da tarefa de modelagem e elicitação.
- Implementação de uma ferramenta colaborativa de suporte ao método.
- Aplicação do método em ambientes organizacionais com processos de negócio reais.
- Avaliação de resultados para aquisição de evidências e conclusões sobre o método e sua aplicabilidade na área de Processos de Negócio.

1.6 Organização da Dissertação

Esta dissertação está organizada em 8 capítulos. O primeiro, Introdução, apresenta a pesquisa em linhas gerais. O Capítulo 2 define os principais conceitos sobre a área de modelagem de processos e aprofunda os problemas abordados. Os Capítulos 3 e 4 tratam de uma levantamento bibliográfico sobre as áreas utilizadas para a proposta de pesquisa, Group Storytelling e Mineração de Texto.

Os capítulos seguintes tratam, em seqüência, da proposta do método (Capítulo 5), sua implementação (Capítulo 6) e os estudos de casos e avaliação dos resultados (Capítulo 7).

Este documento se encerra com o Capítulo 8, onde são discutidas as conclusões resultantes da pesquisa e os trabalhos futuros. Em seguida temos as referências bibliográficas e os apêndices da dissertação.

Capítulo 2 – Elicitação e Modelagem de Processos de Negócio

2.1.Definição

Modelos de processos de negócio têm sido adotados por organizações em variados contextos, como fonte para análise de melhores práticas, projeto de sistemas de informação e construção de arquiteturas de tecnologia da informação.

Um modelo de processo de negócio é uma representação do conjunto de atividades em uma organização, possuindo uma estrutura que descreve sua ordem lógica e dependência, cujo objetivo é a produção de um resultado desejado (AGUILAR-SAVEN,2002).

A elaboração de modelos de processos de negócios pressupõe a tarefa de elicitação de processos de negócio, que busca adquirir o conhecimento envolvido no processo e tornar possível sua utilização pelo analista, e a modelagem de processos de negócios, que constrói os modelos com base na informação coletada previamente.

A elicitação de processos de negócio, também chamada de levantamento de informações para processos de negócio, pode ser definida como a tarefa de tornar explícito o conhecimento sobre um processo de negócio específico, a partir de uma ou mais fontes de conhecimento, sejam seres humanos, documentos, sistemas de informação, etc.

2.2.O conhecimento na elicitação de processos de negócio

Segundo NONAKA e TAKEUCHI (1997) o conhecimento pode ser classificado em dois tipos: tácito e explícito. O primeiro se encontra na mente de cada indivíduo, fruto de suas experiências pessoais, percepções e personalidade. Já o conhecimento explícito é aquele que se encontra expresso por meio de texto, imagens, palavras, números ou outras linguagens de representação.

Um dos problemas mais relevantes do levantamento de informações é a dificuldade de aquisição do conhecimento de cada pessoa envolvida, já que este se encontra geralmente tácito, disperso e de difícil acesso. A elicitação de processos de negócio, sendo um tipo de levantamento de informações, também sofre do mesmo "problema do conhecimento", já que o conhecimento envolvido em um processo de negócio também pode ser dividido em tácito e explícito (KALPIC & BERNUS, 2006).

DAVENPORT e PRUSAK (1998) definem o conhecimento como "o conjunto de experiências, valores, informações contextuais e *insights* de especialistas que fornece uma forma para avaliar e incorporar experiências e informação; é originado e se aplica na mente das pessoas; nas organizações, geralmente encontra-se espalhado não só em documentos ou repositórios, mas também nas rotinas organizacionais, processos, práticas e normas". O conhecimento é uma forma de desenvolvimento cultural e depende de fatores externos ao indivíduo, podendo ser adquirido e compartilhado de diversas maneiras.

O uso de modelos de processos de negócio pode ser considerado parte de uma abordagem para gestão de conhecimento, já que estes representam elementos integrantes do capital intelectual da organização (BOSILJ-VUKSIC, 2006). Os modelos descrevem o conhecimento prático existente em suas atividades e executados pelos seus membros, registrando o que realmente ocorre no funcionamento da organização.

Os autores ainda discutem um ciclo de transformação do conhecimento por meio da chamada Espiral do Conhecimento (vide Figura 1), que representa a transição entre o conhecimento tácito e explícito nas seguintes formas: Externalização, Combinação, Internalização e Socialização.

Figura 1 – Espiral do Conhecimento de NONAKA & TAKEUCHI (1997)

A externalização é a transformação do conhecimento tácito em explícito. Ela pode acontecer por meio da representação formal de conceitos, hipóteses ou modelos.

A combinação é a transformação de conhecimento explícito em explícito, relacionando e categorizando conhecimento armazenado em diversos repositórios (como arquivos, documentos, etc.), seja por meio de sistemas de informação ou até mesmo conversas e reuniões. Este processo gera novas conclusões, premissas e idéias, ou seja, novos tipos de conhecimento explícito.

Já a internalização é a transformação de conhecimento explícito em tácito. Seja por meio de manuais, documentos ou livros, esta transformação constitui o processo de aprendizado individual, o "know-how" técnico e compartilhado.

A socialização é definida como a transformação do conhecimento tácito em tácito. As observações, as práticas e as imitações são exemplos de como o conhecimento tácito é compartilhado pela socialização.

Embora todas as transformações da espiral do conhecimento sejam relevantes para a elicitação de processos de negócio, podemos destacar a externalização como uma das transformações mais importantes e críticas, pois está diretamente relacionada com a tarefa posterior que é a modelagem de processos.

Outra transformação digna de nota no contexto do presente trabalho é a socialização, onde ocorre a interação entre o conhecimento tácito de cada indivíduo com seu colega de trabalho. Para a tarefa de elicitação de processos, é importante frisar que raramente uma única pessoa conhece o processo de negócio em sua totalidade ou com o grau de detalhe adequado.

As outras duas transformações, a internalização e a combinação, também são relevantes neste contexto, já que refletem o aprendizado de técnicas e modos de trabalho e o uso de sistemas de informação para organizar informações de forma mais prática e útil para a organização, ou seja, são os usos dos modelos de processos.

2.3.A Abordagem Típica de Elicitação de Processos de Negócio

A abordagem típica para levantamento de informações e posterior modelagem de processos envolve um analista que conduz entrevistas com os executores das tarefas, documenta estas informações em formato de texto e depois representa o conhecimento documentado através de modelos. O analista submete estes modelos aos participantes do processo, que os refinam para conseguir a máxima fidelidade à realidade. Observa-se, no entanto, que este método apresenta alguns problemas.

O primeiro deles é que geralmente as perguntas elaboradas pelo analista tendem a guiar o usuário (ALVAREZ, 2004). Assim, a ênfase pode ser colocada em aspectos que o analista supõe serem importantes, fazendo com que a coleta de informações sobre o processo identificado dependa em grande parte da habilidade do analista responsável.

Um segundo problema, também relatado na engenharia de requisitos de software (PACHECO, TOVAR, 2007), pode ocorrer durante a escolha dos participantes do processo a ser modelado que relatam os processos: as pessoas possuem perspectivas distintas devido aos papéis que assumem, e têm experiências e percepções diferentes (BECKER, DELFMANN, KNACKSTEDT *et al.*, 2002 *apud* BECKER, DELFMANN, DREILING *et al.*, 2004). O resultado tenderá a ser um modelo que pode não contemplar as características reais do processo, mas sim a opinião de um grupo restrito de pessoas envolvidas com as atividades.

Outro problema reside na organização das informações levantadas. As informações devem ser concisas, evitando-se o excesso de documentação, principalmente de informações redundantes ou fora de escopo para o processo-alvo (GOGUEN & LINDE, 1993). Este obstáculo está relacionado com os problemas de comunicação entre

analistas - responsáveis por fazer os levantamentos - e entre os modeladores - responsáveis por interpretar as informações levantadas e construir os modelos de processos.

2.4.Extração automática de processos de negócio (*Process Mining*)

Outra abordagem bastante utilizada para elicitação e modelagem de processos de negócio é a extração automática, a partir de arquivos de "log" de sistemas de informação, denominada "Mineração de Processos" ou "*Process Mining*" (AALST & WEIJTERS, 2005). A mineração de processos tem como objetivo a extração de atividades pertencentes aos processos de negócio que estão sendo diretamente apoiadas por sistemas de informação, tais como SAP e ERP, e a geração posterior (usualmente utilizando redes de Petri como notação) de um modelo representativo do processo.

Um dos maiores benefícios da mineração de processos é o fato de ser uma abordagem totalmente computacional, sem ter que lidar com as pessoas envolvidas em tarefas relacionadas ao processo que se deseja elicitar. Desta forma, os problemas relacionados ao conhecimento, pontos de vista divergentes, ausência de consenso, o viés do entrevistador interferindo nas informações adquiridas, típicos da técnica de entrevistas, são evitados em prol de um enfoque em torno de sistemas de informação.

No entanto, esta abordagem possui duas desvantagens intrínsecas: a primeira se refere à origem das informações sobre o processo. Um sistema de informação que automatiza atividades necessariamente embute um modelo de processos a ser seguido, seja este um modelo "ad-hoc" ou o trabalho de analistas de processos que foi transferido para uma ferramenta computacional. No entanto, não existe garantia que ele reflita totalmente a realidade, já que está baseado nas informações dos sistemas, e não no conhecimento das pessoas envolvidas diretamente com as atividades. Embora, em alguns casos, esta característica seja uma vantagem, no caso de processos onde existe um alto grau de atividades humanas, podem ser geradas incoerências e até mesmo a inconsistência do modelo extraído.

A segunda desvantagem reside no fator humano, geralmente negligenciado no caso da mineração de processos. Atividades que envolvam decisões, ou que possuam um fator humano externo ao sistema, simplesmente não serão capturadas.

2.5.Trabalhos relacionados

Na área de Elicitação de Requisitos de Software, diversas publicações foram encontradas tratando dos mesmos problemas da elicitação de processos de negócio. Pode-se perceber que as publicações da área se dividem em dois grupos distintos: artigos que possuem um enfoque empírico, baseados em dados de revisões bibliográficos ou até mesmo revisões sistemáticas da literatura, e artigos que possuem um enfoque qualitativo, comparando técnicas de modo a extrair evidências gerais sobre a área ou linhas gerais a serem seguidas para a seleção de técnicas de elicitação.

No primeiro grupo, se destaca uma revisão sistemática da literatura (DAVIS, DIESTE & HICKEY, 2006), que possui como objetivo a busca por evidências empíricas sobre a área de elicitação de requisitos. Os artigos selecionados foram filtrados utilizando três regras simples:

1) Artigos devem realizar um estudo empírico genuíno, e não serem apenas a validação de uma abordagem específica,

2) As técnicas utilizadas devem ser simples e individuais, sem suporte por computador ou utilizadas em grupos,

3) O foco de estudo das publicações deve ser a comparação de diferentes técnicas, uma mesma técnica em diferentes situações ou a aquisição de informações de especialistas da área.

Com estas condições, reduziu-se o escopo das publicações a serem analisadas para apenas aquelas que podem apresentar evidências empíricas mais evidentes sobre a área.

Uma importante conclusão final deste estudo foi que a técnica de entrevistas, anteriormente descrita também como a "forma tradicional de elicitação de processos de negócio", foi considerada a técnica com maior grau de eficiência no maior número de situações.

Esta evidência se comprovou mesmo num segundo estudo (DIESTE, LOPEZ & RAMOS, 2008), realizado como atualização da revisão sistemática anterior, o que demonstra a força do fato descoberto pela pesquisa. Desta forma, a técnica de entrevistas se comprova como "ponto de referência" para avaliação de novas técnicas para elicitação de processos de negócio, de modo que devemos examiná-la

minuciosamente para descobrir seus pontos fortes e fracos, e criar propostas a partir da análise de seus resultados.

Outros autores, como DAVEY & COPE (2008), utilizando técnicas menos formais de revisão bibliográfica, também chegaram à mesma conclusão sobre as técnicas de entrevistas. Indo um pouco mais além, eles definiram a problemática da elicitação de requisitos em duas causas principais: dificuldade de comunicação entre as pessoas e mudanças das necessidades da organização com o passar do tempo.

Para a elicitação de processos de negócio, o enfoque na dificuldade de comunicação entre os indivíduos envolvidos em um processo é bastante claro, sendo mais um ponto em comum entre as duas áreas.

Denominado "aspecto conversacional", o enfoque na comunicação entre pessoas se torna de tal importância para a área, que é considerado como um grupo à parte de outras técnicas por ZHANG (2007), separando-o das técnicas que envolvem apenas observação ou uso de documentação pré-existente.

Esta é uma característica importante da elicitação de requisitos, possuindo relação com todo o processo de Engenharia de Requisitos, devido à sua importância e influência na qualidade do resultado final. A importância da elicitação é ainda mais destacada por HICKEY & DAVIS (2004), que identificam a atividade como a mais abrangente dentre as atividades da Engenharia de Requisitos, como na Figura 2.

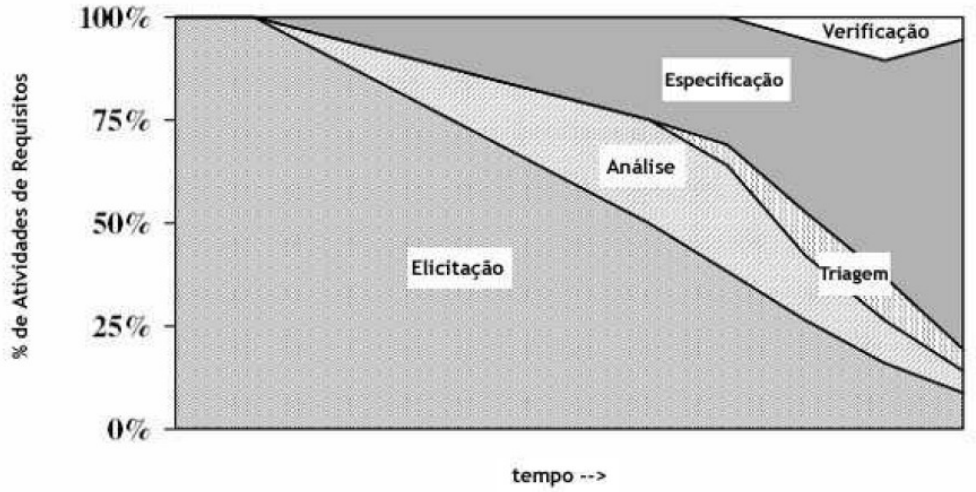


Figura 2 – Atividades da Engenharia de Requisitos, extraída de HICKEY & DAVIS (2008)

Com a forte evidência sobre a importância do aspecto conversacional, do ponto de referência da técnica de entrevistas para fins de comparação e as semelhanças entre os problemas e necessidades das áreas de Elicitação de Requisitos e Elicitação de

Processos de Negócio, resta saber como selecionar uma técnica da área e como avaliar sua viabilidade em uma situação específica.

TSUMAKI & TAMAI (2005) propõem uma comparação entre técnicas de elicitação de requisitos, instituindo linhas gerais para sua seleção. Duas características ou "dimensões" foram levadas em conta para o estudo. A primeira delas, operação, escreve como o processo de elicitação de requisitos é conduzido, se uma análise sistemática e estruturada é utilizada (categoria "estática") ou o foco é o seu contexto dinâmico, visando compreender os procedimentos e mudanças com o passar do tempo (categoria "dinâmica"). A segunda dimensão, objeto, aborda o grau de conhecimento prévio e domínio sobre o conhecimento alvo. Se o objeto é estável e conhecido (categoria "fechado"), ele pode ser elicitado por formulários e outras formas mais estruturadas. Em contrapartida, se o objeto é desconhecido e mutável, ele é considerado parte da categoria "aberto".

A Figura 3 mostra o quadro comparativo de várias técnicas de elicitação conhecidas e seu posicionamento em relação aos critérios utilizados.

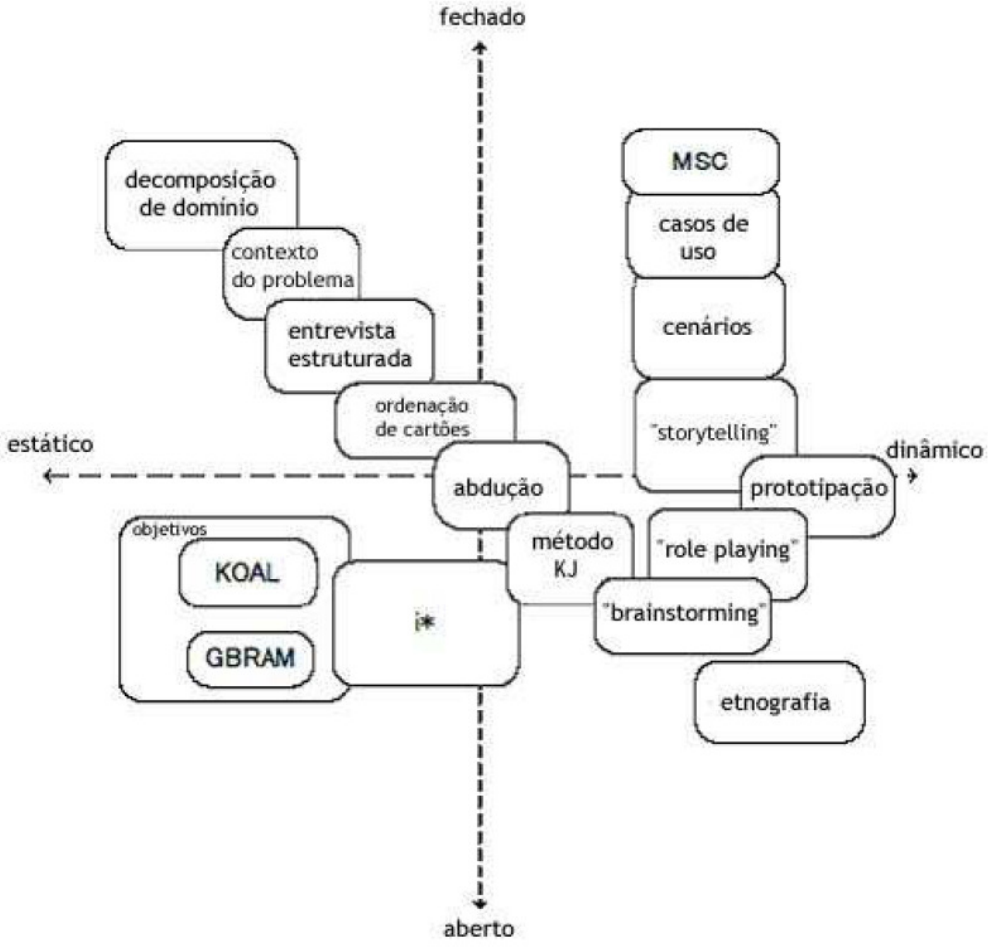


Figura 3 – Diagrama de Técnicas de Elicitação de Requisitos (TSUMAKI & TAMAI, 2005)

Dentre as técnicas analisadas, podemos realizar uma rápida comparação com a área de interesse da dissertação, a Elicitação de Processos de Negócio. Sendo um processo algo definido como parte das tarefas de uma organização, geralmente se conhece qual processo será modelo, nem que seja ao menos o seu nome ou tema. Está é uma diferença crucial da área em relação à Elicitação de Requisitos de Software, onde muitas vezes se conhece pouco sobre o que será realmente elicitado. Outro aspecto da área de processos de negócio é a inabilidade de se executar uma análise estruturada, já que temos que levar em conta pontos de vista diferentes, documentação de diversos departamentos e diferentes dimensões do conhecimento, que muitas vezes podem gerar conflitos e incoerências.

Com base nisso, podemos levar em conta os aspectos “dinâmico” e “fechado” da Figura 3 para avaliação de técnicas de Elicitação de Processos. As técnicas do quadrante

referente a estas duas dimensões são os cenários, os casos de uso, a *Message Sequence Chart(MSC)* e o *storytelling*.

Portanto, as técnicas que chamam a atenção por serem aplicáveis à área de Processos de Negócio são o *storytelling* e os cenários, já que as outras, tais como MSC e casos de uso, são voltadas demasiadamente ao desenvolvimento de software e notações específicas, para sua fácil adaptação.

A técnica de *storytelling* para requisitos de software (LAPORTI, BORGES, BRAGANHOLO, 2009) é classificada pelo framework como possuidora de uma forte característica dinâmica e um leve aspecto fechado (TSUMAKI & TAMAI, 2005). A dimensão dinâmica se deve à liberdade de expressão que devem ter os contadores de uma história e a capacidade de adaptação a mudanças. Já a dimensão fechada pode ser justificada pela necessidade de se conhecer o tema da história a ser narrada.

O uso de cenários se mostra eficiente para a elicitação de requisitos de software (ROLLAND, SOUVEYET, ACHOUR, 1998), o que pode indicar uma aplicação promissora para a Elicitação de Processos de Negócio, devido às semelhanças com a área de requisitos. É possível criar analogias, por exemplo, entre partes de um cenário e os fluxos principais e alternativos de instâncias de um processo.

No próximo capítulo, a abordagem de *storytelling* será descrita, principalmente a sua versão colaborativa, a técnica de Group Storytelling.

Capítulo 3 – Group Storytelling

3.1.Introdução

Contar histórias é um dos costumes mais antigos da humanidade e uma das formas mais naturais de intercâmbio de conhecimento. Mesmo antes da existência da palavra escrita, povos antigos já utilizavam narrativas para descrever eventos e preservar tradições, sendo parte da natureza humana o uso da narrativa para comunicar idéias. O ato de narrar histórias não se resume, no entanto, apenas a compartilhar conhecimento. Ele também serve para atividades de aprendizagem, formas de entretenimento e para provocar reações e emoções diversas em um grupo de pessoas (VALLE, PRINZ & BORGES, 2002).

Por ser uma prática inerente ao ser humano, seja em uma reunião de empresa ou em uma conversa informal, o uso de narrativas se aproxima bastante do conhecimento tácito do narrador, contendo dentro de si os elementos emocionais da pessoa, além de seus pontos de vista e características de sua personalidade. Desta forma, podemos considerar que a narrativa é uma das formas mais naturais de se tornar explícito o conhecimento tácito, sendo neste caso, uma atividade de externalização, conforme foi visto na espiral de conhecimento de Nonaka (NONAKA & TAKEUCHI, 1997).

3.2.As histórias

Uma história pode ser contada de várias formas. Um único indivíduo pode ser o narrador ou um grupo de pessoas podem trocar experiências sobre um determinado acontecimento ou assunto. O ato de contar histórias também pode variar em relação ao meio de comunicação, desde uma história contada pessoalmente, contos e romances em forma de texto e até mesmo fóruns na Internet.

As histórias contadas em grupo geralmente envolvem indivíduos, distribuídos ou em um mesmo local, que contribuem para criar uma história em conjunto, de forma síncrona ou assíncrona, utilizando diferentes mídias e meios de comunicação para tal (VALLE et al., 2003). Desta forma, o conhecimento gerado no final do processo, presente na narrativa em si, é a combinação do conhecimento tácito de cada um dos participantes, sendo formado por seus pontos de vista, impressões, opiniões e intuições sobre o tema da história.

ALVAREZ (2002), realizando um estudo em entrevistas de elicitação de conhecimento e com base na análise de discurso, aponta uma evidência que parece reforçar o uso de narrativas colaborativas. Foi percebido, durante as entrevistas utilizadas pelo estudo, um "padrão de contar histórias" por parte dos entrevistados. Este padrão, rico em conhecimento e natural para as pessoas, enfrenta no entanto, resistência por parte do analista, que orienta a entrevista de acordo com o modelo de perguntas e respostas, tão comum na elicitação de processos tradicional.

Uma técnica promissora, que foi utilizada para Gestão de Conhecimento (PERRET, BORGES & SANTORO, 2004), Modelagem de Processos de Negócio (SANTORO, BORGES & PINO, 2008) e Elicitação de Requisitos (LAPORTI, BORGES, BRAGANHOLO, 2009) é o Group Storytelling. A sua principal vantagem é a construção de narrativas colaborativamente e será descrita na próxima sub-seção.

3.3.A técnica Group Storytelling

Group storytelling (Em português "Construção de Narrativas Colaborativas") é a técnica de construção de histórias na qual "mais de uma pessoa contribui, síncrona ou assincronamente, localmente ou de maneira distribuída, em vários pontos do processo, através de diversas mídias" (VALLE; PRINZ; BORGES, 2002). A técnica, e sua capacidade de representar o conhecimento tácito de diversos indivíduos, endereçam diretamente as dificuldades da Elicitação de Processos de Negócio, já que ambas possuem objetivos similares, voltados a representar o conhecimento de pessoas, geralmente voltado a um objetivo, assunto ou tema da história.

A aplicação da técnica geralmente se dá em um formato assíncrono e distribuído, em geral utilizando-se de ferramentas computacionais para permitir o acesso facilitado aos participantes, embora o uso de ferramentas não seja obrigatório. Desta forma podemos eliminar a necessidade de termos todos os participantes em um mesmo lugar ao mesmo

tempo. Devido à dinâmica das organizações, principalmente se forem de grande porte e multinacionais, estes fatores facilitam a aplicação da técnica, podendo ser estendida a vários departamentos e filiais localizadas em cidades, estados, e até mesmo países diferentes.

Outro ponto forte é a vantagem de possibilitar inclusões à história a qualquer momento, após "insights" sobre o tema ou em um momento propício da escolha dos participantes, ao invés de obrigar a criar partes da narrativa em horários pré-definidos (PERRET, 2004).

O Group Storytelling não se restringe aos fatos narrados em ordem cronológica, contendo outros elementos como os personagens, que participam dos acontecimentos da história. Sendo partes inseparáveis das narrativas, os personagens possuem características próprias, como objetivos e planos, à parte dos fatos presentes na história. Outro elemento importante é o contexto da história. Diversos tipos de informação contextual podem ser encontrados em narrativas, desde os locais onde ocorrem os acontecimentos descritos, até informações sobre tempo decorrido, estado emocional dos personagens e causas ou conseqüências das ações realizadas.

Desta forma, a narrativa é definida como uma estrutura conceitual que fornece eventos e elementos auxiliares, de modo a descrever acontecimentos e o contexto em que ocorrem. Esta definição condiz com a necessidade da Elicitação de Processos de Negócio, onde se busca adquirir conhecimento sobre o que ocorre em um conjunto específico de atividades ou tarefas em uma organização. A informação contextual, como por exemplo os lugares onde a história ocorre ou informações sobre os personagens envolvidos, também se faz necessária, pois permite a compreensão das condições em que ocorrem os elementos do processo.

3.4.As histórias e a memória organizacional

As narrativas colaborativas representam não apenas fatos, mas também parte da experiência pessoal e visão pessoal do narrador sobre os acontecimentos registrados. Com a aplicação sucessiva da técnica e a geração de várias narrativas, estas podem ser armazenadas para serem utilizadas pela organização. Este "repositório de histórias", rico em conhecimento possui características, que o aproximam de outro conceito importante da Gestão de Conhecimento (SMITH, 2001), a Memória Organizacional.

De forma intuitiva, a Memória Organizacional (MO) pode ser definida como "como o registro de dados, informações e conhecimento úteis para a organização, de forma que ela possa reaproveitá-los". Alguns autores apresentam definições formais, com uma visão fortemente tecnológica, como (ABECKER & DECKER, 1999), enunciando que a Memória Organizacional "integra técnicas básicas em um sistema computacional que continuamente coleta, atualiza e estrutura conhecimento e informação, e os provê em diferentes atividades operacionais de forma sensível ao contexto, intencionada e ativa".

Embora as definições sejam adequadas, principalmente para a área de sistemas de informação, a restrição do conceito de Memória Organizacional a técnicas computacionais ou ao simples registro de informações ignora o conhecimento proveniente dos indivíduos, partes integrantes de uma organização.

Outra definição mais adequada a esta visão é a de ACKERMAN (1994). Segundo o autor, "a Memória Organizacional pode estar retida na cultura organizacional, nas transformações organizacionais, na ecologia organizacional, nos arquivos externos, em manuais corporativos, nas bases de dados, nas histórias organizacionais e nos indivíduos".

Neste caso, as narrativas e as informações complementares da técnica Group Storytelling podem ser consideradas como uma instância da Memória Organizacional e como representação do conhecimento que flui pela organização como um todo. O processo de criação das histórias será detalhado na próxima seção.

3.5.O Processo do Group Storytelling

O processo de construção de narrativas colaborativas, estimulando a liberdade de expressão como foco principal, envolve idéias, sentimentos, conceitos e pontos de vista bastante pessoais de cada participante. A dinâmica original se divide em três fases: Construção, onde se cria a história propriamente dita, Redação, quando os eventos da narrativa são transformados em texto e Conclusão, com o armazenamento da história final para uso posterior (PERRET, 2004).

Em cada uma das etapas, papéis específicos são definidos para as pessoas envolvidas, com base em seu perfil e conhecimento sobre os fatos relatados. Todos eles derivam do papel básico de Colaborador, que define um participante genérico do processo de contar histórias.

Os elementos envolvidos na narrativa não se restringem apenas aos fragmentos da história, ligados por uma linha de tempo e denominados Eventos. Outros elementos também são típicos do processo como Comentários, que são formas de crítica e sugestão às partes da história pelos participantes e Documentos, que representam fontes de conhecimento relacionadas ao assunto abordado, porém externas ao fluxo da narrativa.

Embora o consenso e uma história final satisfatória para os envolvidos e que reflita o processo a ser modelado sejam os objetivos finais da técnica, muitas vezes, pode ocorrer a divergência de opiniões e pontos de vista sobre determinada parte do processo, o que pode ser considerado como conseqüência direta da própria riqueza de conhecimento da dinâmica de Group Storytelling.

Para este fim, dois elementos adicionais da narrativa podem ser utilizados para amenizar os possíveis conflitos de idéias e solucionar entraves para a criação de uma história coerente e representativa da visão de todos os envolvidos: Votações e Versões.

Nas próximas subseções, será feita uma descrição de detalhada das fases envolvidas na dinâmica de Group Storytelling, além da definição dos elementos envolvidos na narrativa..

3.6.Fases do Group Storytelling

Quatro papéis devem ser exercidos pelos participantes na realização da dinâmica. O mais geral deles é o de Colaborador, participando da história como narrador. Outros papéis podem ser acumulados pelos usuários: Comentador, Redator e Moderador.

O Colaborador é o papel mais básico, e representa as atribuições gerais de todos os participantes. O Colaborador pode contribuir livremente na inclusão de informações e das discussões da equipe, incluindo eventos, comentários, personagens e documentos.

Além de ser um Colaborador, o Moderador pode alterar ou excluir trechos da história construída, além de atribuir papéis aos outros participantes. Espera-se que o moderador seja alguém que domine a tema a ser tratado na história e possua a habilidade de lidar com possíveis conflitos de opiniões que possuam surgir durante o processo de criação de uma narrativa colaborativa.

O redator é o papel responsável pela redação final da história. Possui a atribuição de gerar um texto único, uniforme, ou seja, aperfeiçoar o texto gerado a partir do fluxo de

eventos da história. É vital um domínio sobre o idioma no qual a história está escrita, além de habilidade em escrita e narrativa.

O Comentador é o papel que, após a geração do texto final da história, é responsável por indicar os elementos tácitos do conhecimento expressos pelos participantes. Este papel exige um conhecimento profundo sobre gestão de conhecimento e as variáveis envolvidas como os dois tipos de conhecimento, além da externalização do conhecimento tácito.

Existem três fases principais na aplicação da técnica de Group Storytelling (vide Figura 4). Na primeira fase, ocorre a **Construção** das histórias em si. Os usuários podem registrar eventos, de acordo com sua experiência, além criar novos personagens, comentários e inserir documentos relevantes para a história. A criação de novos elementos permite que os membros do grupo se lembrem de outros fatos, o que possibilita a criação de mais eventos ou outros elementos relacionados.

Este ciclo de criação de elementos, com novos fatos sendo relembrados pelos participantes, é parte fundamental da construção da narrativa colaborativa, já que permite maior quantidade de informações do que em uma história contada por uma única pessoa. Os eventos registrados seguem um fluxo, representando a linha temporal da narrativa. Operações, como mudar de lugar dois eventos, por exemplo, podem ser realizadas, caso seja necessário. A história se conclui quando o fluxo construído é considerado como satisfatório pelos participantes, e mais nenhuma modificação é realizada.

A segunda fase, chamada de **Redação,** se inicia após a conclusão da Construção. Um ou mais participantes são selecionados para o papel de Redator e a tarefa de escrever a história definitiva começa. Os eventos são concatenados e transformados em um texto, respeitando-se a coerência e as relações lógicas presentes.

Por fim, a última consiste na **Conclusão** da história, quando esta fica disponível para consulta e para que alguns dos participantes sejam escolhidos pelo Moderador para servirem como Comentadores e registrar os trechos onde exista conhecimento tácito externalizado.

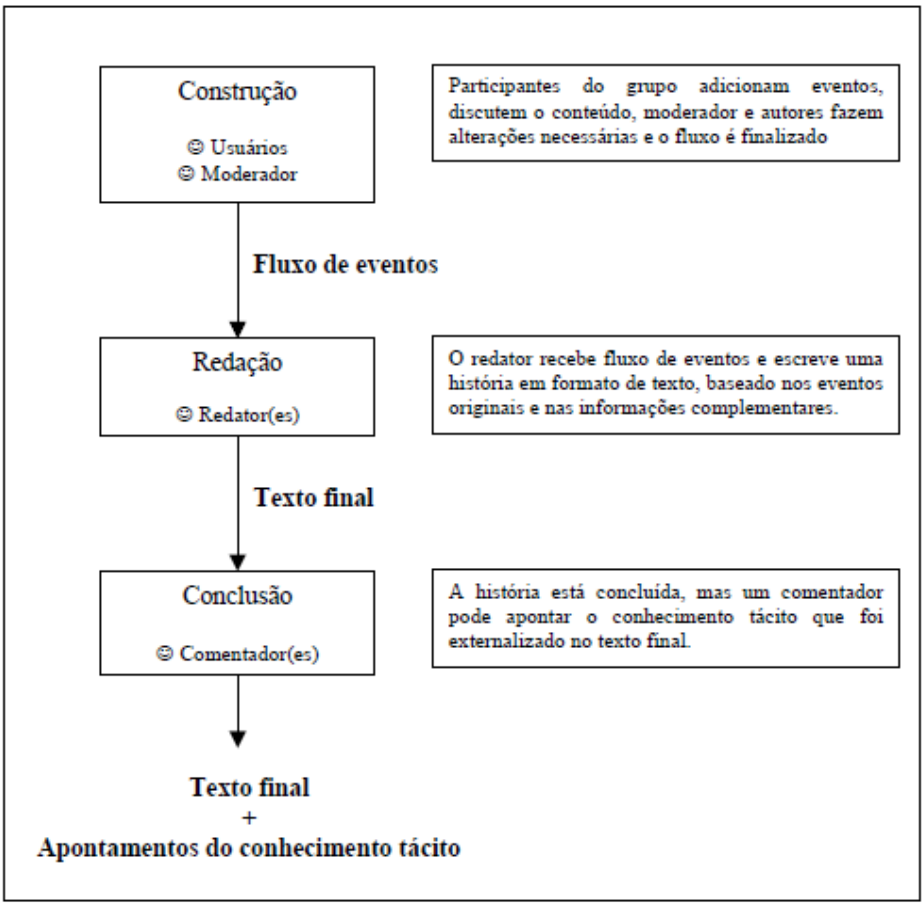


Figura 4 – Estrutura da dinâmica de Group Storytelling (PERRET, 2004)

3.7.Elementos da Narrativa

Em geral, as narrativas contêm elementos comuns, independente de assunto ou contexto. Estes são chamados de "elementos" de uma história por serem comuns a todas elas. Na técnica de Group Storytelling, por causa do enfoque na construção colaborativa de narrativas, suas partes possuem características específicas, de modo a se adequarem à dinâmica. Elas são: Eventos, Comentários, Documentos, Votações e Versões (PERRET, 2004).

3.7.1. Eventos

Na construção das narrativas em grupo, o foco principal está nos eventos que compõem a história (SANTORO *et al.*, 2005). Os eventos são formados por relatos de ocorrências reais, com o adicional da visão ou ponto de vista do narrador. Os fatos representados nos eventos, geralmente descritos em linguagem natural, formam as parcelas do

conhecimento presente nas histórias, possuindo relacionamentos explícitos entre si de modo a formar o fluxo de eventos da narrativa, já que uma história não é apenas um conjunto de eventos, possuindo também uma linha temporal que ordena os fatos descritos.

Uma característica das narrativas em grupo é a discussão entre os participantes sobre a linha temporal dos eventos. Alguns exemplos são questões sobre um evento específico, sobre quais os seus eventos anteriores e posteriores, além do número de eventos necessário para descrever uma parte da história. Esta dinâmica traz a necessidade de mecanismo para manipulação do fluxo de eventos, tornando possível a troca de lugar de eventos na linha temporal, a união de dois ou mais eventos e a fragmentação de um evento em duas ou mais partes.

3.7.2. Personagens

Os personagens são partes integrantes de qualquer narrativa, representando os participantes ou agentes nos eventos da história, envolvidos nos fatos relatados e interagindo com os acontecimentos. A definição dos personagens de uma história não se restringe aos indivíduos, como à primeira vista pode parecer, mas podem ser objetos inanimados, animais, ou organizações, por exemplo.

3.7.3. Documentos

Documentos são fontes auxiliares de conhecimento sobre o tema abordado nas histórias. Geralmente são planilhas, relatórios, normas, regulamentos e qualquer outra forma de informação que possa ser relevante para a história a ser construída. Os documentos adquirem uma importância maior quando o assunto a ser tratado é muito específico, sendo uma fonte adicional de referência sobre a narrativa.

3.7.4. Comentários

Comentários são utilizados com os eventos da história como forma de contribuição que outros usuários podem fazer aos eventos registrados por um participante. Eles são partes integrantes do processo de construção da narrativa colaborativa, permitindo uma maior comunicação entre os narradores, além de registrar as opiniões sobre o processo de formação da história como um todo.

3.7.5. Votações

Votações são ferramentas para resolver visões divergentes sobre uma determinada parte da história. Embora a técnica Group Storytelling busque o consenso entre os participantes, de modo a formar a narrativa mais fiel possível a todos os pontos de vista combinados, existe a possibilidade de termos uma diferença de opiniões muito forte, que deve ser tratada de forma adequada.

As votações permitem que todos os envolvidos possam decidir, de forma democrática, qual dos rumos em discussão será utilizado na história final. Caso haja uma decisão pela maioria de votos, a história segue o rumo escolhido. Em caso de empate, é necessária a criação de uma nova versão, a partir do evento em disputa.

3.7.6. Versões

As versões são criadas quando existe mais de um caminho a ser seguido na história. Isso ocorre quando um consenso não é atingido e um evento se torna o ponto divergente da história. A nova versão será criada a partir do evento divergente, gerando na prática uma bifurcação no fluxo de eventos da história.

O principal objetivo das versões é permitir que o fluxo da história possa refletir esta ausência de consenso sobre uma de suas partes sem afetar sua consistência, nem criar uma divisão entre os participantes impossível de se solucionar.

No entanto, para evitar a criação de um grande número de versões, o que acarretaria, na prática, em várias histórias ao invés de uma única história representativa, costuma-se limitar o número de versões a apenas duas (sendo uma a história original e a segunda sua "bifurcação", a partir de um de seus eventos).

3.8.Group Storytelling aplicada a Processos de Negócio

Originalmente, a técnica de Group Storytelling foi utilizada para Gestão do Conhecimento de maneira geral. A aplicação da técnica para a Elicitação de Processos de Negócio ainda é recente e poucos trabalhos relacionados foram encontrados na literatura. SANTORO, BORGES & PINO (2008) propõem o uso de Group Storytelling para elicitação de processos de negócio, no qual os atores de um processo descrevem de

forma colaborativa como suas atividades são executadas por meio de narrativas. Uma figura ilustrativa do método proposto pelos autores pode ser vista na figura 5.

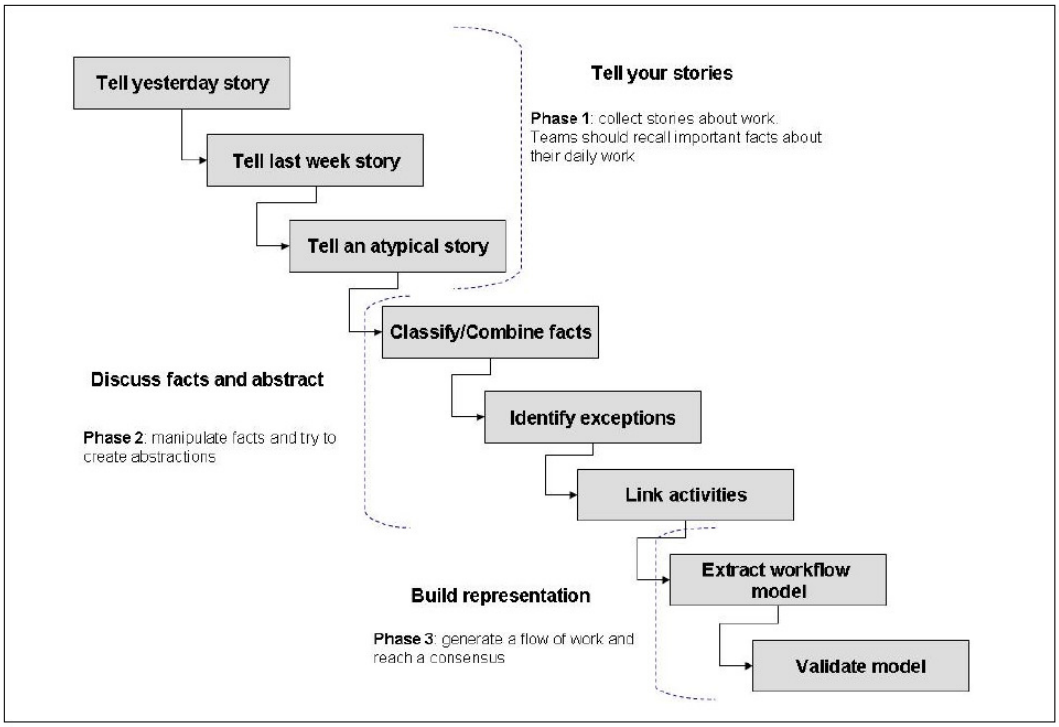


Figura 5 – Método para Group Storytelling aplicado à Elicitação de Processos (SANTORO, BORGES & PINO, 2008)

A proposta se diferencia um pouco das três fases descritas por PERRET (2004), sendo composta por uma fase inicial, onde fatos concretos são contados pelos participantes, passando a uma etapa posterior onde temos a descoberta de abstrações e classificação de fatos, terminando com o produto final do método: um modelo de workflow. O enfoque voltado a um objetivo mais específico do que as tarefas de gestão de conhecimento em geral, trazem à tona a necessidade de algumas modificações no método.

Outra diferença clara se faz entre os eventos de uma história e os eventos de um processo. Embora à primeira vista uma relação direta seja aparente, um evento de uma história pode possuir diversos eventos do processo a ser elicitado, apenas um evento ou até mesmo nenhum evento, visto que os participantes possuem liberdade sobre o que escrever no seu conteúdo.

A principal mudança se concentra nos papéis atribuídos aos participantes, agora com três papéis distintos: Narrador, Facilitador e Modelador. Narradores são pessoas, que participam no processo de construção de histórias, registrando fatos e informações, tornando-as explícitas através da narrativa. Já o facilitador é um profissional experiente

na construção de narrativas, que fornece suporte aos narradores, auxiliando-os a conectar os fatos e estimulando a coerência na narrativa, embora não interfira diretamente na expressão do conhecimento. O modelador é um analista de processos que deve refinar o modelo gerado pelos eventos das histórias e gerar um modelo gráfico com base nas abstrações retiradas delas.

Embora a técnica de Group Storytelling possua a grande vantagem da riqueza de conhecimento, além de ser assíncrona e distribuída, facilitando a reunião dos pontos de vista relevantes para o processo, a narrativa em linguagem natural possui algumas dificuldades inerentes à livre expressão de conhecimento pelos narradores. Estes obstáculos serão discutidos na próxima seção, além de possíveis soluções para eles.

3.9.Dificuldades da manipulação do conhecimento adquirido

Devido à livre expressão do Group Storytelling e ao estilo pessoal de cada indivíduo durante a composição de sua narrativa, algumas dificuldades surgem para o uso posterior da informação presente nas histórias. Elas se devem justamente a uma das principais vantagens da técnica sobre outras formas tradicionais de Elicitação de Conhecimento: a liberdade de expressão concedida ao participante. Transformar o texto livre obtido como resultado da aplicação da técnica de Group Storytelling em um modelo de processo, que tradicionalmente é um gráfico do tipo fluxo de trabalho não é trivial. Isso envolve a interpretação do conteúdo do texto e identificação dos principais elementos que devem compor este modelo, tais como atividades, papéis e fluxos.

Um obstáculo comum deste tipo de narrativa é o uso de gírias, siglas e outros símbolos e expressões características de uma linguagem peculiar aos membros de uma equipe ou departamento e com significado de difícil compreensão para pessoas de fora. Outro empecilho é o surgimento de novas ambigüidades e informações irrelevantes para o processo em questão, em vista da liberdade de expressão dos narradores em contar fatos ocorridos. Este problema envolve um tópico de pesquisa bastante relevante e interdisciplinar, denominado de Processamento de Linguagem Natural.

Podemos definir a linguagem natural como a forma comum de expressar idéias, utilizada pelo homem. As informações são representadas e transmitidas por meio dela, em suas diferentes formas, como escrita, fala, etc.

Na forma escrita, ela se mostra extremamente rica em informações e bastante confortável para o ser humano. No entanto, estas duas características, embora sejam benéficas ao texto a ser composto e à posterior transmissão de informações por meio deste, cria alguns problemas do ponto de vista computacional, exatamente pela forma "livre" de expressão e pela carência de um padrão previsível. Estes problemas são provocados, principalmente, pelo uso de frases e períodos ambíguos e complexos, que geram uma grande dificuldade de compreensão e posterior tradução ou processamento do texto tanto para seres humanos quanto para computadores (MITAMURA & NYBERG, 2001).

Algumas iniciativas podem ser utilizadas de modo a tentar reduzir este problema. O uso de linguagem controlada, limitando a forma de expressão de conhecimento por parte do narrador é uma possível solução, no entanto, esta abordagem vai de encontro ao principal benefício da técnica: a livre expressão de conhecimento.

Outras soluções nesta mesma linha incluem o uso de descrições formais para as histórias, tanto de seu contexto como dos eventos, utilizando linguagens baseadas em Lógica Matemática como o Cálculo de Eventos (KOWALSKI & SERGOT, 1986) e o Cálculo de Situações (MILLER & SHANAHAN, 1994). Porém, isto gera um novo problema: a obrigatoriedade de uso de uma linguagem formal, o que também limitaria a expressão de conhecimento.

O obstáculo, em resumo, reside na relação entre a riqueza de conhecimento presente na livre expressão de conhecimento e a dificuldade de interpretação e aquisição deste mesmo conhecimento presente nas histórias. Uma área bastante ativa de pesquisa em Ciência da Computação se mostra promissora ao se aproximar de uma solução para este problema, sem interferir no processo de externalização de conhecimento dos participantes. Ela se denomina Mineração de Texto e será tratada no próximo capítulo.

Capítulo 4 – Mineração de Textos

4.1.Definição

A Mineração de Textos pode ser definida, de forma geral, como "o processo intensivo de conhecimento no qual um usuário interage com uma coleção de documentos utilizando uma série de ferramentas de análise" (FELDMAN & SANGER, 2007).

Originalmente concebida como uma aplicação de Mineração de Dados em textos, a Mineração de Textos (MT) possui um objetivo similar, ou seja, a extração de conhecimento útil a partir de fontes de dados. No entanto, ela se diferencia da Mineração de Dados devido a suas fontes de dados serem não estruturadas (coleções de documentos) e seus padrões encontrados serem formados por relações entre os diversos elementos textuais não-estruturados presentes em documentos de texto, entre documentos, ou até mesmo entre coleções de documentos.

HEARST (2003) define a mineração de texto como "o processo de descobrir computacionalmente novas informações, previamente desconhecidas, pela extração automática de informação de diferentes recursos de texto". Um elemento chave para este processo seria a utilização das informações extraídas, para se formar novos fatos, novas hipóteses e conclusões previamente desconhecidas.

A analogia de "mineração" é criada desta forma, já que é similar ao processo de mineração tradicional, onde se busca, em meio a pedras e terra, encontrar metais preciosos (HEARST, 1999). No caso da MT, o "metal precioso" a ser extraído é a informação ou conhecimento previamente desconhecido.

4.2.Mineração de Textos e Áreas relacionadas

HOTHO, NÜRNBERGER & PAASS (2005) definem a MT como um "método verdadeiramente interdisciplinar", utilizando conceitos e técnicas de outras áreas do conhecimento como Recuperação de Informação, Extração de Informação e Processamento da Linguagem Natural. Isto se deve ao enfoque principal no documento de texto e pelo fato da Mineração de Texto como área de pesquisa ser relativamente recente, quando comparada com áreas tradicionais da Ciência da Computação como a Mineração de Dados.

Apesar da aparente similaridade entre as áreas correlatas, existem diferenças claras entre elas. A Mineração de Dados tem como objetivo a extração de padrões de formas estruturadas de dados como bancos de dados relacionais, por exemplo, enquanto o objetivo geral da área de Processamento da Linguagem Natural (PLN) é alcançar um maior grau de compreensão da linguagem natural pelo uso de computadores (KODRATOFF, 1999). Sendo uma área bastante antiga da Ciência da Computação, suas técnicas são variadas e incluem desde a simples manipulação de strings de texto até o processamento automático de documentos e análise lingüística de seus elementos constituintes.

Uma diferença menos óbvia está nas áreas de Recuperação de Informação (RI) e Extração de Informação (EI). A EI pode ser definida como “uma forma limitada de compreensão de textos, onde já se conhece o tipo de informação que se deseja extrair” (HOTHO, NÜRNBERGER, PAASS, 2005). Na prática, ela limita o escopo do processamento do grande número de elementos presentes em um texto, de acordo com um objetivo definido. A EI também é considerada como uma das áreas mais proeminentes nas aplicações de Mineração de Texto (KAISER & MIKSCH, 2005). Já a RI lida com a representação, armazenamento, organização e o acesso à itens de informação, visando fornecer ao usuário acesso fácil à informação desejada. (BAEZA-YATES & RIBEIRO-NETO, 1999).

Segundo GAIZAUSKAS & WILKS (1998), existe uma diferença clara entre as duas áreas: a RI trata da descoberta de documentos relevantes em conjuntos de documentos enquanto a EI lida com a extração de informação relevante de documentos.

A Mineração de Texto, portanto, se diferencia das outras áreas correlatas com seu enfoque de descobrir conhecimento previamente desconhecido a partir de padrões, fatos e informações extraídas de texto. A utilização de métodos e técnicas de áreas correlatas é vital para MT, visando atingir dois objetivos principais: estruturar o texto e extrair padrões. No caso da estrutura do texto, a principal área envolvida é o Processamento da Linguagem Natural (PLN), com décadas de pesquisas em busca de uma forma computacional de compreender a estrutura gramatical e semântica da linguagem. Para a extração de padrões, duas áreas se destacam: A Mineração de Dados e a Extração de Informação. No caso da primeira, os algoritmos aplicados em base de dados têm sido adaptados para utilização de grandes conjuntos de textos, utilizando suas estruturas como guias para a mineração, de forma semelhante às aplicações tradicionais em bancos de dados. Já as técnicas de Extração de Informação frequentemente são utilizadas quando existe um objetivo específico a ser alcançado. Conhecendo-se o tipo de conhecimento ou estrutura a ser extraída, o uso de métodos de EI para a Mineração de Texto permite que se extraia conhecimento novo de forma robusta e eficiente.

No decorrer deste capítulo serão descritos os conceitos básicos da Mineração de Texto, o processo genérico e suas fases, além de métodos de duas áreas correlatas utilizadas nesta dissertação: a Extração de Informação e o Processamento da Linguagem Natural.

4.3.Conceitos Básicos de Mineração de Texto

Os conceitos mais básicos da área de Mineração de Texto são os Documentos, as Coleções de Documentos, os Modelos Representativos dos Documentos, o Conhecimento Externo ou Domínio de uma aplicação.

4.4.3.1. Documentos

Um documento pode ser definido de maneira informal como uma “unidade discreta de texto que geralmente, mas não necessariamente, pertence a uma coleção de documentos” (FELDMAN & SANGER, 2007). Geralmente, mas não necessariamente, a unidade de texto corresponde a algum tipo de documento do “mundo real” como um relatório, memorando, artigo, etc. Um documento pode ser definido como uma unidade de texto que se refere a um contexto específico e inclui um período e um assunto específicos.

Os documentos podem ser classificados de acordo com sua estrutura. Embora geralmente se defina texto como "dados não-estruturados", os documentos em si podem ser definidos em duas categorias: "semi-estruturados" ou "fracamente estruturados". Documentos baseados em texto, como páginas da Web, arquivos PDF e DOC, por exemplo, possuem uma estrutura interna e modelos a serem obedecidos, como "*tags*" XML e informações sobre fontes e paginação. Estes são geralmente denominados documentos semi-estruturados. Documentos de texto livre, compostos apenas de texto e com pouca informação adicional em sua estrutura, são denominados de "formato livre" ou "fracamente estruturados". Mesmo em um documento de formato livre, a mineração pode ser realizada, já que elementos como a pontuação, uso de caixa alta e caixa baixa, numeração e caracteres especiais, por exemplo, formam uma estrutura rica em detalhes que pode ser utilizada pelas técnicas de PLN.

As histórias, por seus eventos serem descritos em forma de texto livre, se caracterizam como documentos fracamente estruturados, embora as informações complementares a ela (documentos, personagens e o próprio fluxo temporal onde os eventos estão inseridos) possam servir como complemento.

Como a MT tem um enfoque voltado aos elementos presentes nos documentos e não nos documentos em si, geralmente existe um equilíbrio entre dois objetivos: o primeiro é encontrar um conjunto eficiente e útil de elementos que representem com fidelidade semântica os documentos e o segundo objetivo é descobrir elementos que sejam computacionalmente viáveis e úteis para a descoberta de padrões e de conhecimento. Este conjunto de elementos forma o chamado "modelo representativo do documento", que será visto na próxima subseção.

4.4.3.2. Modelos Representativos dos Documentos

De modo a tornar possível a extração de conhecimento, o documento deve ser definido em relação aos elementos que o compõem e quais relações e sentidos diferentes estes elementos possuem entre si. Este conjunto de elementos é denominado modelo representativo do documento, que descreve os diferentes elementos que um tipo abstrato de documento possui. Os elementos do modelo representativo servem como guia para diversas tarefas como a classificação de conjuntos de documentos por estrutura, por exemplo.

O modelo representativo de um documento servirá como base para a criação de uma estrutura, a partir do conteúdo do documento, de modo que o torne passível de aplicação de algoritmos que levem em consideração algo mais do que conjuntos de caracteres.

Em geral, os elementos que compõem o modelo representativo de um documento são elementos de texto, como palavras, frases e parágrafos. Isto se aplica principalmente nos documentos fracamente estruturados. Já em documentos semi-estruturados, podemos ter *tags* HTML ou XML e até mesmo padrões de estruturas complexas, como linhas e fórmulas em planilhas eletrônicas ou código escrito em uma linguagem de programação, o que possibilita utilizar um modelo representativo mais complexo. Desta forma podemos dizer que os elementos presentes nos modelos representam o documento em si, a nível de aplicação..

A vasta gama de possíveis modelos representativos de documentos traz algumas dificuldades, principalmente em relação à quantidade de elementos que pode possuir. Existem dois problemas distintos neste caso, denominados alta dimensionalidade dos elementos (*high feature dimensionality*) e esparsidade de elementos (*feature sparsity*).

A alta dimensionalidade se caracteriza pelo excesso de elementos em um modelo representativo, levando a uma abundância de padrões a serem minerados e uma dificuldade ainda maior em extrair conhecimento novo pelo processo de mineração. Este tipo de problema é muito mais frequente e grave na mineração de textos do que na mineração de dados em geral. Isto se deve à linguagem natural, que geralmente possui uma gama maior de conceitos representados do que registros de um banco de dados.

A esparsidade de elementos ocorre quando apenas uma pequena parte dos elementos do modelo representativo é encontrada em cada documento da coleção. Isto gera uma dificuldade ainda maior em encontrar padrões, devido à carência de variáveis úteis para relacioná-los.

Geralmente, embora diversas classificações de elementos possam ser utilizados para representar um documento, na maioria das aplicações em geral (e para os objetivos desta dissertação) quatro tipos básicos são utilizados: Caracteres, Palavras, Termos e Conceitos (FELDMAN & SANGER, 2007).

4.3.2.1. Caracteres

A representação por caracteres é a forma mais básica de se representar um documento, já que estes são os elementos mais básicos, os quais compõem outros elementos de nível semântico mais alto como palavras, termos e conceitos. Uma representação deste tipo pode incluir todos os caracteres do documento ou um subconjunto específico deles. Geralmente, a informação posicional de cada caractere também é registrada, já que apenas os caracteres que compõem um texto seriam de pouca valia para a mineração de textos. Embora seja a representação mais básica de um documento, pode-se afirmar que é considerada a mais completa representação de um documento do mundo real, já que retrata exatamente o conteúdo do documento em si.

4.3.2.2. Palavras

Palavras específicas selecionadas diretamente de um texto podem ser apontadas como o nível mais básico em termos de riqueza semântica. Elas se diferenciam dos caracteres por possuírem um significado próprio. Por causa disto, as palavras de um documento são, às vezes, denominadas o "espaço de elementos nativo" (YANG & PEDERSEN, 1997) deste.

Embora a representação de todos os elementos de um documento em nível de palavras seja possível, geralmente alguma forma de otimização é utilizada, de modo a extrair um subconjunto mais útil.

Dentre as técnicas que utilizam a representação de um documento por suas palavras, podemos citar a filtragem de *stopwords*, que são palavras, símbolos e números sem sentido semântico ou utilidade prática para a aplicação de mineração desejada. Outra importante aplicação utilizando este elemento é a freqüência de cada palavra em uma coleção de texto, que permite identificar quais palavras são mais representativas dentre todos os documentos.

4.3.2.3. Termos

Embora possuam diferentes definições na literatura, podemos definir os termos como palavras ou frases de várias palavras, identificadas com base em padrões morfo-sintáticos e selecionadas diretamente do corpo de um documento por meio de um

método de extração de termos (FELDMAN, FRESKO, KINAR *et al.*, 1998). Portanto, podemos definir o conjunto de termos de um documento como as palavras e expressões específicas que são representativas nele, para uma determinada aplicação. Este tipo de representação pode ser construído de diversas maneiras, desde a utilização de um repositório de termos específicos (Ex: Termos pertencentes a um domínio específico, como Medicina, Informática,etc.) até mesmo extraindo-se termos que reflitam padrões gramaticais, como Sujeito, Verbos e Advérbios.

Desta forma, o conjunto de termos de um documento utilizado por uma aplicação sempre será um subconjunto de todos os termos do mesmo. Este elemento é uma abstração em um nível mais alto do que as palavras, possuindo um nível semântico ainda maior. O objetivo é uma representação reduzida do documento, com uma riqueza semântica maior do que a simples extração de palavras.

4.3.2.4. Conceitos

Conceitos são elementos de uma riqueza semântica mais alta do que os termos, representando abstrações que podem se refletir em fragmentos de texto (BLAKE & PRATT, 2001). Eles podem ser gerados para um documento de diversas maneiras, geralmente pela análise de seu conteúdo ou por uma fonte de conhecimento externa. Este processo pode ocorrer manualmente, ou utilizando algum método de classificação mais complexo.

A representação de documentos através de conceitos não está restrita a palavras ou partes do texto. Um conceito denota um significado por si só, que pode estar presente de diversas formas em um mesmo documento. Desta forma, aplicações, que utilizam este elemento como base, podem identificar conceitos que não estão presentes de forma literal no corpo de documento. Por exemplo, se aplicarmos uma técnica deste tipo a uma coleção de documentos sobre carros esporte, pode-se extrair o conceito “automóvel”, mesmo que esta palavra não esteja presente em nenhum documento da coleção.

A maioria dos métodos que envolvem conceitos possui algum grau de interação com uma fonte externa de conhecimento, que pode variar de acordo com o método utilizado. Esta pode ser formada por um conjunto de documentos representativos de um domínio (HOLZINGER, GEIERHOFER, MÖDRITSCHER *et al*, 2008)., uma ontologia (HOTHO, STAAB, STUMME, 2004) ou, até mesmo pelo *feedback* com o usuário (HUANG & MITCHELL, 2006).

Dos quatro elementos básicos descritos nesta seção, os conceitos e termos são os mais úteis para as aplicações de Mineração de Texto, já que possuem uma semântica própria associada aos seus elementos, a parte do conteúdo de cada documento. No entanto, principalmente no caso de Conceitos, há uma dificuldade muito maior de implementá-los do que no caso de aplicações que utilizam os elementos mais básicos, como caracteres e palavras, por exemplo.

4.4.3.3. Coleções de Documentos

A definição mais elementar de uma coleção de documentos é "qualquer grupo de documentos baseados em texto" (FELDMAN & SANGER, 2007). Na prática, as aplicações de Mineração de Texto estão interessadas em grandes coleções, de onde podem ser extraídas relações e padrões relevantes sobre um assunto específico.

As Coleções podem ser classificadas como estáticas, cujo número de documentos não varia, e os documentos não sofrem modificações com o passar do tempo, ou dinâmicas, cujos documentos sofrem alterações e pode ocorrer a adição de novos documentos com o passar do tempo. Esta divisão é importante para o projeto de aplicações de software, já que define diversos requisitos para o uso de técnicas de mineração. Em cenários em que a coleção é dinâmica, o aspecto de desempenho (no que se refere ao custo computacional) da técnica de mineração de texto é tipicamente mais crítico.

O conteúdo de cada documento presente em uma coleção possui diferentes tipos de informação, geralmente referente a um assunto específico. Este mesmo documento pode estar presente em mais de uma coleção e pode existir em diferentes coleções ao mesmo tempo, de acordo com o critério de cada coleção. Um exemplo seria um artigo de jornal sobre "o investimento do governo em tecnologia para a educação", que poderia estar presente em coleções classificadas pelos temas Governo, Educação, Tecnologia, Computação, Política, Economia, dentre outros.

Várias técnicas visam as coleções para mapear termos e conceitos presentes em seus documentos e extrair conhecimento dos padrões encontrados. Com o advento da Web, este campo de pesquisa em MT se estendeu inclusive para a Web, sendo utilizada para objetivos mais complexos como a descoberta de relacionamentos em potencial dentro de redes sociais (SANTOS, EVSUKOFF, LIMA *et al.*, 2009)

4.4.3.4. Domínios e Conhecimento Externo

As fontes externas de conhecimento possuem um papel importante na Mineração de Texto, principalmente no uso de elementos como termos e conceitos para encontrar padrões. Mais importante ainda, as fontes de conhecimento de um domínio específico (ex: Aviação, Economia, etc...) permitem trabalhar eficientemente com conceitos presentes em um texto, ampliando as possibilidades de aplicações da área.

Na área de processos de negócio e gestão de conhecimento, é bastante comum o uso de um vocabulário próprio, principalmente relacionados a áreas específicas ou à cultura organizacional de uma organização específica. As fontes externas de conhecimentos podem facilitar a tarefa de mineração, criando uma forma de acessar o vocabulário específico de um grupo ou organização, possivelmente extraindo conhecimento de melhor qualidade.

De forma geral, estas fontes se traduzem de alguma forma estruturada, desde tesauri (TSUJII, ANANIADOU, 2005) até formas mais complexas como ontologias de domínio (GARCIA, FERRAZ, PINTO, 2006). Na prática, a principal função destas fontes é facilitar a descoberta do conhecimento, permitindo identificar parcelas de texto relevantes dentre todos os documentos disponíveis.

4.4.Processo de mineração de textos

De forma geral, um sistema de Mineração de Textos utiliza como entrada uma fonte de texto (documentos ou coleções de documentos) e gera uma série de saídas, geralmente envolvendo alguma forma estruturada de conhecimento e informação útil para o usuário (padrões encontrados, visualizações, tendências, gráficos, etc...). Embora seja uma definição simplista, ela ilustra perfeitamente o universo externo à aplicação. Definindo a fonte de texto inicial para processamento e o resultado desejado, a forma geral do processo de mineração de textos é conforme a Figura 6.

Figura 6 – Processo geral de Mineração de Texto

Em um nível mais detalhado, podemos perceber que o processo se inicia com o texto de entrada, que passa por etapas ou fases específicas, como pré-processamento, até alcançar o resultado desejado, a extração de padrões e de conhecimento, como saída. Neste processo, o usuário também pode interagir com o sistema de forma iterativa, armazenando resultados anteriores e utilizando suas percepções e anotações para utilizar futuras operações de mineração pelo mesmo processo descrito, muitas vezes com textos de entrada diferentes ou algoritmos específicos para cada objetivo..

De forma mais detalhada, as fases podem ser definidas como:

4.4.1. Pré-Processamento

A fase de pré-processamento inclui todas as rotinas, algoritmos e operações destinadas a preparar o dado de forma "bruta" para que ele seja processado pelas fases posteriores. O enfoque é na explicitação ou geração de modelos representativos dos documentos de entrada.

Os métodos presentes nesta fase incluem desde ordenação de documentos por data, separação de fragmentos texto por parágrafos, frases, palavras, termos ou conceitos, até a classificação automática ou manual de partes do texto e a geração de novos documentos ou estruturas de dados a partir de seu conteúdo. O uso de técnicas de Processamento da Linguagem Natural é bastante comum nesta fase, por possuir um papel fundamental na criação de uma estrutura a partir dos elementos de texto, auxiliando na aplicação de técnicas em fases posteriores.

Nesta seção, são apresentadas algumas de técnicas de PLN mais utilizadas em mineração de textos.

4.4.1.1. Filtragem por elementos básicos e *stopwords*

Em geral, estas são as primeiras técnicas a serem utilizadas, diretamente no texto original. Elas incluem a separação do texto em elementos básicos segundo o modelo de representação adotado e a eliminação de elementos considerados irrelevantes (*stop words*).

A descoberta das *stopwords* pode otimizar as atividades posteriores de Mineração de Texto. Definidas como palavras freqüentes demais ou que não são pertinentes para uma aplicação (DOLAMIC & SAVOY, 2009), os sistemas de mineração que utilizam este tipo de filtragem possuem listas pré-definidas de acordo com a linguagem utilizada nos documentos, ou usam algum tipo de técnica para detectar as palavras repetitivas ou irrelevantes do corpo dos textos a serem processados.

4.4.1.2. Ocorrência de Palavras

Documentos costumam conter centenas ou até milhares de palavras. As coleções de documentos, portanto, contem um número ainda maior de palavras se contabilizarmos todos os documentos pertencentes a alguma coleção específica, por exemplo. Esta é uma técnica muito utilizada para a indexação e classificação de documentos e para a descoberta automática de palavras-chaves representativas de um domínio.

De forma simplificada, um dos padrões mais básicos de se encontrar em qualquer documento ou coleção é a freqüência de cada palavra, ou seja, a ocorrência de cada palavra específica em um documento ou conjunto de documentos. Uma das medidas mais utilizadas para representar a relevância de uma palavra em uma coleção de documentos alvo é a TF/IDF (*Term Frequency/Inverse Document Frequency*), (ROBERTSON, 2004), que é calculada como a seguir.

Em sua forma mais comum, TF/IDF é definida pela seguinte fórmula (FELDMAN & SANGER, 2007):

$$TF\text{-}IDF(w, d) = TermFreq(w, d) \cdot \log(N / DocFreq(w))$$

Sendo uma palavra (**w**) em um documento (**d**), **TermFreq(w,d)** é a frequência desta palavra no documento, a variável **N**, o número de documentos presentes na coleção de documentos e **DocFreq(w)** a quantidade de documentos da coleção que contém a palavra **w**.

4.4.1.3. Resolução de Anáforas e Co-referência

Anáfora é o fenômeno lingüístico onde um trecho do discurso (referência) remete a um referente, uma palavra que já ter sido mencionada, que ainda será mencionada no decorrer do texto ou que é externo ao discurso (CHAVES & RINO, 2008). A posição no texto alvo do referente e a existência explícita da referência são informações importantes para a aplicação de técnicas de resolução de anáforas.

Portanto, de início, deve-se localizar onde se encontra a anáfora e a referência propriamente dita para, num momento posterior, realizar a sua substituição pela palavra apropriada. O objetivo da resolução de anáfora é simplificar o texto para a posterior aplicação de algoritmos de mineração de texto.

Diversos algoritmos foram propostos na literatura, utilizando uma variedade de técnicas para localização e resolução de anáforas. Um dos mais famosos é o algoritmo de Hobbs (HOBBS, 1986), que realiza a comparação de cada sintagma nominal com a anáfora, estabelecendo uma série de regras para detectar qual deles seria a referência.

4.4.1.4. POS-Tagging (Part-Of-Speech Tagging)

Part-of-Speech Tagging é o nome dado às técnicas de classificação de palavras baseadas em seu contexto, onde o contexto de uma palavra reflete o papel exercido por ela em uma frase (CUTTING, KUBIEC, PEDERSEN *et al.*, 1992). O conjunto (*tagset*) mais comum de etiquetas de classificação inclui sete classes (*tags*): Artigo, Substantivo, Verbo, Adjetivo, Preposição, Numeral e Nome Próprio.

A maioria dos outros "*tagsets*" utiliza variações deste conjunto básico, de acordo com objetivos específicos. Em geral, os algoritmos deste tipo também usam um *corpus* de documentos, conjunto de documentos com palavras e suas respectivas "*tags"*, para treinamento, ou um conjunto de regras para guiar a tarefa de classificar automaticamente as palavras de um texto não-classificado.

Estes corpora (sing. corpus) são específicos para cada idioma, e variam por tema e assuntos dos textos que os compõem. Também existem corpora representativos de todo um idioma, como o Brown Corpus (BROWN, 1967) para a língua inglesa e o Corpus MAC-MORPHO (ALUISIO, PINHEIRO, MANFRIM, 2004) para a língua portuguesa, dentre outros.

Algumas aplicações mais complexas de *POS-tagging* englobam outras técnicas para refinar seus resultados, como a redução de palavras aos seus radicais (*stemming*) (ORENGO & HUYCK, 2001).

Um dos tipos mais comuns de "taggers" ou etiquetadores morfossintáticos é o N-Gram Tagger (BIRD, KLEIN, LOPER, 2009). Um N-Gram Tagger é um algoritmo, treinado em um corpus anotado de treinamento ou algum outro tipo de fonte de conhecimento sobre palavras e suas classificações, que possibilita a classificação morfossintática de palavras de um texto. O "N-Gram" se refere ao contexto lingüístico que será utilizado para treinamento e classificação, ou seja, o número de palavras que será utilizado como padrão de comparação. Por exemplo: Unigram Tagger (uma palavra), Bigram Tagger (duas palavras), Trigram Tagger (três palavras) (BIRD, KLEIN, LOPER, 2009)).

Sendo um dos POS-tagger mais básicos, outras técnicas utilizam este algoritmo e refinam seus resultados. Um dos exemplos mais conhecidos é o Brill-Tagger (BRILL, 1992), que aplica aprendizado baseado em transformações (TBL – *Transformation based learning*) para gerar um conjunto de regras e prever qual a classificação mais correta, de acordo com o corpus de treinamento, de uma palavra desconhecida pelo etiquetador.

4.4.1.5. Parsing Sintático

As técnicas de parsing sintático realizam uma análise sintática aprofundada nas frases do texto, com base em uma teoria gramatical (FELDMAN & SANGER, 2007). Elas se dividem em dois tipos, de acordo com a teoria utilizada: gramáticas de constituintes e gramáticas de dependência.

Gramáticas de constituintes representam a estrutura das frases de um texto como uma seqüência de frases, construídas recursivamente a partir de seus elementos sintáticos. A maior parte destas gramáticas estabelece distinções entre diferentes tipos de sintagmas, como sintagmas nominais, sintagmas verbais, sintagmas preposicionais, sintagmas adjetivos, dentre outros. Portanto, cada frase pode conter um número de constituintes que varia entre zero até o total de palavras da frase, de acordo com as regras estabelecidas pela gramática. Além disto, a estrutura sintática entre frases também é verificada, permitindo um maior grau de detalhe na análise realizada. Por exemplo, um sintagma nominal próximo de um sintagma verbal, pode ser considerado como um

possível sujeito, e duas frases conectadas por uma conjunção, poderiam indicar uma possível relação entre elas.

As gramáticas de dependência, por outro lado, não possuem um enfoque nos sintagmas como unidades lingüísticas separadas, focando apenas nas relações diretas entre palavras. Uma típica análise de dependência consiste no uso de uma estrutura de dados contendo diversas palavras e as específicas dependências de cada uma delas. Por exemplo, uma relação de sujeito e objeto direto em uma frase depende do verbo principal, e uma relação de adjetivo depende do substantivo que ele modifica.

De um modo geral, as frases de um texto podem ser extraídas por uma análise de dependência, sendo uma das técnicas mais populares e simples de se utilizar. No entanto, as gramáticas de dependências possuem problemas com construções lingüísticas mais sutis, como conjunções.

4.4.1.6. Shallow Parsing

A análise eficiente e precisa dos componentes lingüísticos de um texto, embora teoricamente possível, possui um custo muito alto em processamento e tempo de execução, além de não ser robusta suficiente para seu uso em larga escala nas aplicações práticas de mineração de texto.

Desta necessidade, surgiu a técnica de "shallow parsing", também conhecida como "chunking" (SANG & BUCHHOLZ, 2000), que sacrifica a profundidade da análise sintática em prol de uma maior velocidade de processamento e robustez das técnicas. Ao invés de realizar a análise completa (também denominada "parsing") do texto, esta técnica trata apenas das partes que são relativamente fáceis e não apresentam ambigüidades nos resultados. Geralmente, simples sintagmas nominais e verbais são extraídos, criando uma estrutura inicial a partir da qual as classificações mais complexas serão formadas. Desta forma, as relações e dependências mais proeminentes e claras são descobertas, enquanto as relações mais ambíguas dos componentes textuais são deixadas de lado.

Para a maioria das aplicações de mineração de textos, o "shallow parsing" é muito eficiente e muitas vezes preferível à análise total das relações gramáticas, principalmente por causa de sua maior velocidade de processamento.

4.4.2. Operações de Mineração

As operações de mineração formam o núcleo de uma aplicação de mineração de textos e têm como enfoque a extração de padrões, a identificação de tendências e a aquisição de novos conhecimentos e informações previamente desconhecidos, a partir do texto pré-processado.

As operações de mineração são realizadas a partir da estruturação de texto resultante da fase anterior (pré-processamento), através da aplicação de algoritmos sobre elementos de texto destacados.

Em geral, não apenas a extração de padrões ou conhecimento é realizada, mas também a comparação entre resultados, análise da distribuição e proporção de conceitos em um documento ou coleção de documentos e a avaliação de quão relevante ou interessante para os objetivos finais do usuário e dos sistemas são os resultados encontrados.

Esta fase é também conhecida pelo nome de "destilação de conhecimento" (TAN, 1999) e pode envolver diversas técnicas voltadas a utilizar o texto pré-processado para a extração de padrões.

Técnicas de Extração de Informação são extremamente úteis nesta fase, principalmente para aplicações com um objetivo específico, quando se sabe qual tipo de conhecimento ou padrões desejados. Embora as técnicas de EI possuam objetivos diferentes da área de Mineração de Texto, elas podem ser utilizadas, como parte de um contexto maior e em combinação com outras áreas, como é o caso da pesquisa desta dissertação.

Para aplicar as técnicas de EI, um texto é definido como um conjunto de entidades e relacionamentos entre entidades. Entidades podem ser definidas como as partes semânticas fundamentais presentes em um texto, podendo denotar pessoas (João), papéis (Vendedor), organizações (Secretaria, Universidade), etc.... Elas seriam a unidade mais básica a ser tratada em um texto. Entidades possuem atributos, que representam o segundo elemento fundamental da EI. Atributos representam características da própria entidade a que pertencem (ex: nome, título, idade, etc...). Entidades podem possuir relacionamentos entre si, definidos como fatos, o terceiro elemento básico da EI, por exemplo Vendedor **vendePara** Cliente, Chefe **chefia** Subordinado, Aluno **solicitaMatriculaPara** Secretaria.

O quarto e último elemento básico da EIsão os eventos, que representam atividades nas quais as entidades participam, por exemplo o evento **Venda**, no qual as entidades Vendedor e Cliente participariam.

Em geral, a aplicação de técnicas de Extração de Informação passa por 4 fases distintas (Figura 7). Na primeira fase, o conteúdo dos documentos de entrada é separado em suas partes fundamentais (palavras, frases e parágrafos), de forma similar à fase de pré-processamento da Mineração de Textos. Outra tarefa comum é a extração de "tokens", que são partes do texto definidas como relevantes para o objetivo da aplicação. A segunda fase, Análise Léxica e Morfológica, possui o objetivo de criar uma estrutura básica, geralmente com base na classificação gramatical de palavras, para a posterior extração de conhecimento. A fase seguinte, Análise Sintática, visa aprimorar os resultados da anterior, com um enfoque em tornar explícitas as relações entre palavras e frases. As três primeiras fases da EI são similares às técnicas da fase de Pré-processamento da Mineração de Textos, possuindo o objetivo comum de criar uma estrutura para aplicação de técnicas computacionais.

A última fase, Análise de Domínio, visa extrair informações a partir dos resultados das fases anteriores através de um *template*, específico para os objetivos desejados. As técnicas geralmente utilizadas para tal fim são a extração de padrões, utilizando expressões regulares, a partir da classificação de texto resultante do trabalho das fases anteriores, de modo a vincular elementos do *template* a padrões semânticos ou gramaticais. Operações como resolução de anáforas, típicas do pré-processamento da Mineração de Texto, são também realizadas nesta fase da EI.

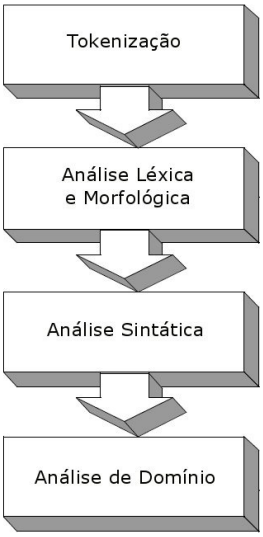


Figura 7 – Processo Geral de um Sistema de Extração de Informação

Segundo este processo, as técnicas de Extração de Informação se baseiam no tipo de informação a ser extraída e no *template* a ser utilizado. Elas são baseadas em algum modelo ou fonte de conhecimento externa específica, de modo a guiar os procedimentos a serem executados.

Geralmente, utiliza-se algum tipo de correspondência entre os elementos da linguagem e os elementos do modelo, de modo a aprimorar a eficácia do processo como um todo.

De acordo com CUNNINGHAM (2007), as técnicas de Extração de Informação se dividem em 5 grupos, que serão descritos a seguir:

- Reconhecimento de Entidades Nomeadas
- Extração de Elemento de Template
- Extração de Relacionamento de Template
- Extração de Template de Cenário
- Extração de Coreferência

4.4.2.1. Reconhecimento de Entidades Nomeadas

O Reconhecimento de Entidades Nomeadas (*Named-Entity Recognition*) é uma das técnicas mais básicas da EI e possui um objetivo simples: a extração de nomes próprios ou previamente conhecidos do texto. Entidades Nomeadas podem ser definidas como

frases que contém nomes de pessoas, organizações ou locais (SANG & DE MEULDER, 2003). Esta definição pode ser estendida para qualquer outro tipo de entidade com nomes próprios. Geralmente se utiliza algum corpus de referência ou fonte de conhecimento. De acordo com o grau de especificidade da aplicação desejada, o corpus de referência pode se tornar bastante dependente de domínio.

A aplicação desta técnica costuma ser utilizada também para restringir o escopo do texto ao objetivo ou domínio específico da aplicação, de modo que as técnicas aplicadas posteriormente enfoquem apenas as partes do texto onde as entidades selecionadas aparecem.

4.4.2.2. Extração de Elemento de Template

O reconhecimento de elemento de template foi proposto originalmente por CHINCHOR (1997). Elementos de um template são independentes de domínio e de cenários, geralmente se referindo a um conceito ou abstração, como pessoa, local, organização, dentre outros. Estes geralmente representam entidades a serem extraídas e formam a base de trabalho para as operações de EI em texto.

Um exemplo simples de um template seria "Pessoa" como um sintagma nominal capitalizado, "Organização" como um sintagma nominal capitalizado contendo palavras como "Ltda.", "Idade" como um numeral seguido da palavra "anos", "Cargo" como um sintagma nominal contendo palavras como "Diretor" ou "Diretora" e "Local" como uma entidade nomeada do tipo localidade. O trecho abaixo mostra como seria uma aplicação simples em um fragmento de texto:

Ex: Nesta sexta-feira, <Pessoa:/"José da Silva"/>, <Cargo:/"Diretor Financeiro"/> da <Organização:/"Silva Ltda."/>, discursou em <Local:/"Brasília"/>. Além dele, também se pronunciaram a <Cargo:/"Diretora de RH"/> da <Organização:/"empresa"/>, <Pessoa:/"Maria da Silva"/>.

Elementos auxiliares também podem ser extraídos como atributos das Entidades. No caso do exemplo acima, seria, por exemplo, a idade de José da Silva.

4.4.2.3. Extração de Relacionamento de Template

As técnicas de extração de relacionamento de template visam encontrar relacionamentos entre entidades previamente extraídas do texto. Relacionamentos possíveis variam de acordo com o contexto do template a ser utilizado e com os objetivos da aplicação.

Por exemplo, um relacionamento "Vínculo", pode ser encontrado em documentos de uma organização em relação a funcionários e às respectivas empresas de que possuem algum vínculo empregatício. Um padrão simples para este relacionamento poderia ser as entidades "Cargo", "Empresa" e "Pessoa" relacionadas por conectivos como "de" e "da" ou a proximidade dos dois elementos.

O exemplo anterior seria incrementado da seguinte forma:

Ex: Nesta sexta-feira, <Vínculo:/<Pessoa:/"José da Silva"/>, <Cargo:/"Diretor Financeiro"/> da <Organização:/"Silva Ltda."/>/>, discursou em <Local:/"Brasília"/>. Além dele, também se pronunciaram a <Vínculo:/<Cargo:/"Diretora de RH"/> da <Organização:/"empresa"/>, <Pessoa:/"Maria da Silva"/>/>.

4.4.2.4. Extração de Co-Referência

A extração de co-referência é bastante útil quando existe um grande número de entidades e relacionamentos extraídos. Esta técnica visa descobrir referências (de diversos tipos, desde pronomes até conjunto de palavras com o mesmo significado) diferentes a uma mesma entidade e agrupar todas as relações e nomes a uma única entidade.

Embora à primeira vista pareça semelhante à técnica de resolução de anáforas do processamento da linguagem natural, ela possui um enfoque mais conceitual do que lingüístico, incluindo outros casos, por exemplo, onde nomes próprios ("João da Silva"), cargos ("O Presidente da Silva Ltda."), títulos ("O Doutor em Informática") e até mesmo abreviações ("Sr. Da Silva"), se referem a uma mesma entidade.

No exemplo desta seção, uma co-referência na frase seria a palavra "empresa", inserida no elemento <Organização:/"empresa"/>, se referindo à <Organização:/"Silva Ltda."/>. Com a substituição, o exemplo ficaria da seguinte forma:

Ex: Nesta sexta-feira, <Vínculo:/<Pessoa:/"José da Silva"/>, <Cargo:/"Diretor Financeiro"/> da <Organização:/"Silva Ltda."/>/>, discursou em <Local:/"Brasília"/>.

Além dele, também se pronunciaram a <Vínculo:/<Cargo:/"Diretora de RH"/> da <Organização:/"Silva Ltda."/>, <Pessoa:/"Maria da Silva"/>/>.

4.4.2.5. Extração de Cenário de Template

Um template de cenário seria voltado para um domínio específico e para os objetivos finais da aplicação, em contraste com as técnicas apresentadas anteriormente. Esta seria a extração dos eventos a partir das partes mais básicas do template, como entidades, atributos e relacionamentos.

Um exemplo seria o evento "estruturar organização", incluindo empregados (entidades "Pessoa" com relacionamento "empregado de" com entidades "Organização") que interagem entre si com um objetivo comum. Este seria parte de um template maior, que poderia incluir outros eventos como "divisão de tarefas", "trabalho em equipe", dentre outros.

Um exemplo simplificado seria a extração de um cenário "PessoaCargoOrganização", extraído da frase de exemplo utilizada anteriormente, a partir do relacionamento Vínculo estabelecido entre os elementos. O resultado seria:

<Cenário:/"PessoaCargoOrganização"/>

<Pessoa:/" José da Silva"/>

<Cargo:/"Diretor Financeiro"/>

<Organização:/"Silva Ltda."/>

<Cenário:/"PessoaCargoOrganização"/>

<Pessoa:/"Maria da Silva"/>

<Cargo:/"Diretora de RH"/>

<Organização:/"Silva Ltda"/>.

4.4.3. Pós-Processamento

Última fase do processo de mineração de textos, o pós-processamento engloba as tarefas de refinar e tornar coerente os resultados da mineração, eliminando a redundância de padrões e informações e aprimorando o grau de clareza, e representar os padrões descobertos na fase anterior, para que o conhecimento extraído seja utilizado de forma eficaz pelo usuário executor do processo.

As técnicas utilizadas nesta fase incluem a agregação de resultados similares ou referentes a um assunto específico (*clustering*), a estruturação das informações encontradas em algum formato ou notação específica, dentre outras.

Um dos principais objetivos desta fase é organizar os resultados e apresentar ao usuário, de uma forma clara e objetiva, para que este possa utilizar de maneira eficiente. Esta parte da aplicação inclui o uso de interfaces gráficas, visualização de resultados, classificações de padrões e informações segundo critérios específicos, navegação interativa pelas coleções de documentos e documentos e anotação em tempo real de conceitos pelos usuários do sistema.

4.4.3.1. Apresentação de Resultados

Uma das partes importantes de um sistema de mineração de texto é a apresentação dos resultados, onde o usuário poderá visualizar os conceitos e padrões presentes nos documentos analisados e interagir com eles de forma a extrair novas informações e conhecimento. Esta apresentação deve incluir funcionalidades de exploração dos diferentes padrões encontrados de forma útil, aplicação de técnicas e algoritmos, além de uma visualização coerente dos resultados da aplicação.

A forma de apresentação deve levar em conta também o perfil do usuário final do sistema. A forma de manipular os resultados da mineração deve ser intuitiva e de fácil compreensão para a pessoa que utilizar a aplicação. A exploração dos diferentes padrões também deve seguir está mesma lógica.

Um elemento importante desta parte da aplicação é a navegação pelos conceitos presentes nos diferentes repositórios e coleções de documentos, especificando suas relações e possibilitando a análise de conceitos individuais, de modo que o usuário possa ter uma visão geral do que está procurando.

Esta parte de um sistema mineração de texto pode levar a soluções ainda mais complexas como linguagens de busca (*query languages*) específicas para uma aplicação (MARTIN, 2005) ou soluções de visualização de conceitos presentes em uma ou mais coleções e documentos (FELDMAN, KLOESGEN, ZILBERSTEIN, 1997).

Para área de processos de negócio, existe um problema similar: a visualização de um processo, de modo que uma pessoa possa compreender a lógica envolvida em suas atividades e obter uma visão concisa sobre o conhecimento expresso em seu modelo (RINDERLE, BOBRIK, REICHERT *et al.*, 2006).

4.5. Considerações sobre Mineração de Texto aplicada aos Processos de Negócio

A aplicação de técnicas de Mineração de Texto em documentos relativos a Processos de Negócio pode auxiliar na interpretação do conhecimento presente na tarefa de Elicitação de Processos de Negócio, provendo suporte ao analista, de modo que ele consiga explorar o conjunto de documentos com base nas técnicas descritas neste capítulo.

Outra vantagem seria o uso de conhecimento do domínio específico do processo, aliado ao suporte do processo de mineração por fontes de conhecimento externo, que criaria uma aproximação entre os conceitos e linguagem utilizados por uma organização e a linguagem particular do analista.

A transformação dos resultados das operações de mineração para uma estrutura de dados pode servir também para a área de Modelagem de Processos de Negócio, já que existem diversas notações e formatos já estabelecidos para a representação de modelos de Processos de Negócio na literatura e no mercado, por exemplo, a notação BPMN (BPMN, 2010).

Mais ainda, os obstáculos de parcialidade no julgamento do analista e grande volume de informação sobre o processo poderiam ter seus efeitos mitigados pela utilização das técnicas de mineração como auxílio para interpretação e acesso ao conhecimento relevante presente na documentação sobre um processo de negócio, por exemplo. No caso da técnica de Group Storytelling, o problema reside na liberdade total de expressão dos narradores sobre fatos ocorridos durante a execução do processo, incluindo possíveis trechos de narrativa pouco relevante para a modelagem de processo como um todo. A aplicação de técnicas de mineração de texto pode auxiliar na separação entre o que representa conhecimento relevante do que não é necessário para a elicitação e

modelagem, além de fornecer suporte ao analista na interpretação da história e documentos associados a ela.

Veremos, no próximo capítulo, como a Mineração de Textos pode ser aplicada para uma solução de Elicitação de Processos de Negócio, utilizando o Group Storytelling como fonte de documentos. O método proposto, denominado *Story Mining*, além de sua implementação e avaliação serão descritos a seguir.

Capítulo 5 – O Método Story Mining

Neste capítulo, o método Story Mining será descrito em detalhe, além das correlações de seus elementos com as áreas descritas nos capítulos anteriores e os problemas da Modelagem e Elicitação de Processos de Negócio abordados pela dissertação.

5.1. Requisitos de um método para elicitação e modelagem de processos

De acordo com todas as questões discutidas nos capítulos anteriores, uma abordagem que traga melhoria significativa para a área de Elicitação de Processos de Negócio e, consequentemente, melhor qualidade para a Modelagem de Processos de Negócio deve possuir os seguintes requisitos principais:

A) Utilizar narrativas, como forma de expressão direta de extração de conhecimento dos participantes do processo.
B) Possuir um enfoque na característica colaborativa da elicitação.
C) Permitir que as pessoas envolvidas no processo de elicitação expressem seus pontos de vista de forma mais livre possível
D) Minimizar os possíveis efeitos negativos da ação do analista na tarefa de elicitação e modelagem como um todo.
E) Utilizar técnicas de extração de informação e mineração de texto, de modo a fornecer suporte à análise das narrativas e geração de versões dos modelos.

Com base nestes cinco requisitos principais, a proposta do método Story Mining (GONÇALVES, SANTORO, BAIÃO, 2009), (GONÇALVES, SANTORO, BAIÃO, 2010b), (GONÇALVES, SANTORO, BAIÃO, 2010c) foi desenvolvida, como uma nova forma de elicitação de processos de negócio. A base do novo método é

proveniente do trabalho de SANTORO, BORGES e PINO (2008), que definiram de forma preliminar um método para elicitação de processos de negócio baseado em Group Storytelling. Na definição inicial, os atores do processo descrevem colaborativamente sua forma de agir através de histórias.

A proposta desta dissertação estende esta idéia inicial, elaborando um método composto de três fases distintas: A primeira é a narrativa colaborativa de fatos concretos pelos participantes e seu posterior armazenamento em um repositório de histórias. A segunda fase é a extração automática de fatos do texto gerado, finalmente culminando com a construção de abstrações e classificações a partir do conhecimento adquirido, tornando possível a geração final de um modelo de workflow.

Desta forma, o método proposto inicialmente foi refinado, de modo que as duas etapas finais (extração automática de fatos e construção de abstrações do processo) fossem suportadas por um processo de mineração de textos. A execução do processo de MT também foi apoiada por uma ferramenta de gerência de workflows . A Figura 8 representa o método proposto.

A maneira informal das narrativas e sua descrição em texto livre proporcionam uma possível dificuldade adicional ao analista, devido à quantidade de informações a ser lida e interpretada. Este risco traz a necessidade de técnicas de suporte ao analista, de modo que este possa visualizar de forma mais clara o conhecimento presente no repositório, sem ter que ler cada uma das histórias. No método Story Mining, esta tarefa será realizada pela aplicação das técnicas de Mineração de Texto, Extração de Informação e Processamento da Linguagem Natural, previamente citadas. A geração de um modelo possível de ser visualizado em ferramentas de modelagem de processos de negócio tem como objetivo também mitigar este problema.

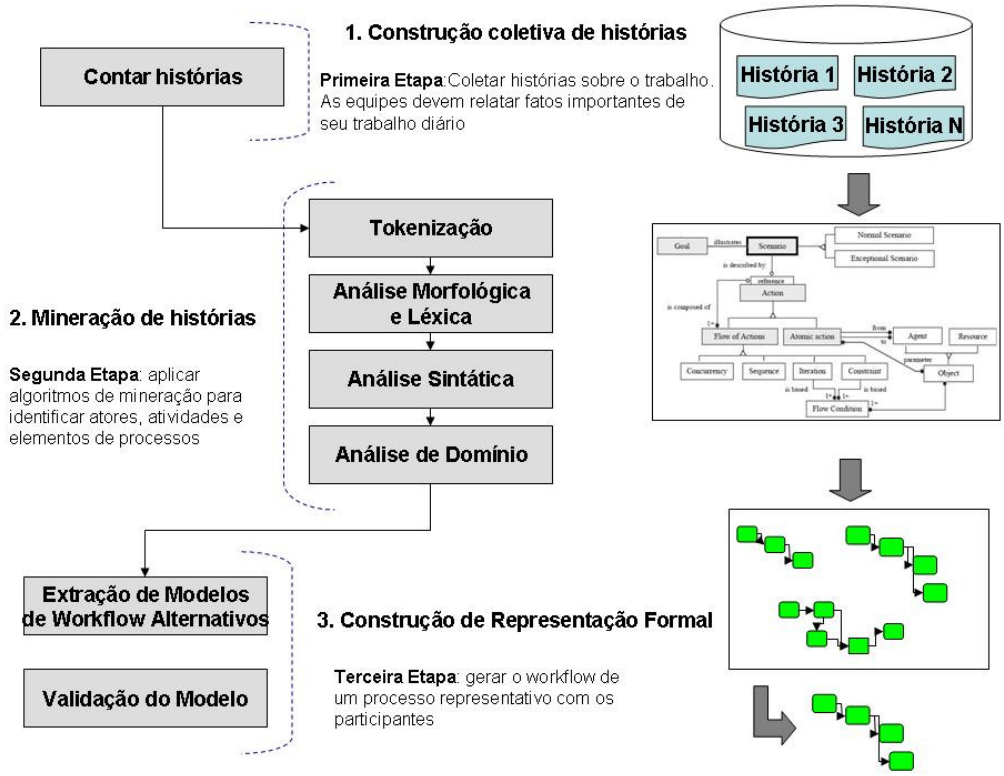


Figura 8 – Método Story Mining

5.2. Papéis envolvidos

Existem três papéis essenciais envolvidos na execução do método: narrador, facilitador e modelador. Os narradores são pessoas que participam do processo e, consequentemente podem explicitar as atividades deste. O facilitador é um profissional experiente que fornece apoio ao processo de narrativa colaborativa, de modo a produzir histórias coerentes. Ambos participam da primeira fase do método, participando da construção coletiva de histórias. Já o modelador participa da terceira fase, sendo um analista de processos que refina o modelo gráfico gerado com base nas abstrações registradas na história.

5.3. Fases do método

As três fases do método, conforme a Figura 8 demonstra são, respectivamente: Construção coletiva de histórias, Mineração de histórias e Construção de representação formal. A primeira fase se resume à construção da narrativa colaborativa, a segunda é a extração de conhecimento em si e a terceiro e última fase conclui o método sendo a análise de resultados e geração de um modelo final, com auxílio dos participantes.

5.4.1. Primeira Fase: Construção coletiva de histórias

Na primeira fase, grupos de narradores são selecionados para participar do processo de elicitação. Eles devem narrar fatos sobre o seu dia-a-dia de trabalho (algo comparável às instâncias do processo a ser modelado). Nesta parte, será utilizada uma ferramenta colaborativa, para prover suporte para a construção de narrativas. As histórias devem ser criadas da forma mais livre possível, de modo que o objetivo da proposta, que é representar os pontos de vista dos participantes de maneira fiel, seja alcançado.

Nesta etapa, se concentra a dinâmica de Group Storytelling descrita no Capítulo 3, com algumas adaptações, principalmente nos papéis envolvidos e nas fases de redação e conclusão.Uma das principais modificações foram os papéis da dinâmica. Enquanto na proposta original existiam os papéis de moderador, narrador, redator e comentador, na aplicação de Group Storytelling para elicitação de processos existem agora o facilitador (com papel semelhante ao moderador) e o narrador (que se mantém da dinâmica original). Estas mudanças se justificam pelo foco no fluxo de eventos como um dos principais resultados finais, em contraste com o enfoque em um texto final que representasse a história, como realizado em sua proposta original (PERRET, 2004).

A importância do fluxo de eventos se deve à sua relação com as atividades de um processo. Uma atividade de um processo é muito mais limitada em escopo do que um evento da dinâmica, sendo restrita a um único ator ou classe de atores. Portanto, um evento de uma história pode conter informações sobre uma ou mais atividades de um processo de negócio, o que não parece óbvio à primeira vista.

Os elementos complementares da história representam o contexto onde o processo ocorre, possuindo relações claras com este. Exemplos seriam os personagens, que são candidatos a se tornaram atores do processo final e os documentos relacionados, que

representam o conhecimento de domínio sobre o processo de negócio a ser elicitado e modelado.

5.4.2. Segunda Fase: Mineração de histórias

A segunda fase se inicia com a análise dos textos produzidos de modo a extrair elementos de um modelo de processo de negócio, como por exemplo, atividades, eventos, fluxos, entradas e saídas. Devido à natureza do texto livre e da livre expressão de conhecimento pelos participantes, a proposta utiliza uma combinação do modelo de histórias colaborativas com um modelo de representação de cenários, de modo a aproximar o texto livre das histórias à representação estruturada de conhecimento. Esta abordagem preserva a riqueza de conhecimento presente nas histórias e, ao mesmo tempo, permite a extração de elementos que constituem um processo de negócio.

O modelo CREWS (ACHOUR, 1998) foi criado como um modelo de representação de cenários, possuindo um enfoque principal em cenários textuais, que podem ser definidos como histórias, descrições de casos de uso ou *scripts*. Ainda de acordo com o ACHOUR (1998), um cenário é "um comportamento limitado a um possível conjunto de interações com um propósito específico, ocorrendo entre diferentes atores". Desta forma, cada cenário é limitado a um objetivo a ser alcançado.

Existe uma similaridade entre cenários e processos de negócio. Ambos são compostos por atores e ações, além de possuírem objetivos específicos. Portanto, o modelo CREWS pode ser utilizado de forma similar a um *template* para aplicação das técnicas de Extração de Informação, para orientar a extração de padrões lingüísticos e a identificação de seus vínculos com elementos de um cenário e, consequentemente, de um processo de negócio.

O *template* foi criado da seguinte forma: os componentes verbais e nominais de cada frase do texto foram descrito em termos de Sintagmas Nominais e Sintagmas Verbais. Conjunções e outros elementos que conecta elementos, que podem indicar relacionamentos entre sujeito e verbo, por exemplo.

O produto final desta etapa, os elementos extraídos e seus relacionamentos servem como base para a construção do proto-modelo da próxima fase. Uma correlação entre os elementos gramaticais e os elementos do modelo de cenários foi realizada:

- Objeto: Sintagmas Nominais que descrevem Agentes (que realizam uma Ação) ou Recursos (que recebem os efeitos de uma Ação)

- Ação: Sintagmas Verbais que descrevem atividades sendo realizadas por um Agente

A partir desta relação dos elementos básicos do modelo de cenários (Objetos e Ações), podemos estabelecer as relações restantes, mais complexas, tais como,

- Fluxos de Ações: Frases que contenham Ações e "palavras-chave" que indiquem o significado de um fluxo específico. No caso do modelo CREWS (ACHOUR, 1998), existem quatro classes de Fluxos de Ações, com alguns exemplos de palavras:
  - Concorrência: "Enquanto", "Simultaneamente", etc...
  - Seqüência: "Após", "Em seguida", etc...
  - Iteração: "Cada", "Para cada", etc...
  - Restrição: "Se", "Nunca", etc...

Estas palavras-chave podem estar armazenadas em um dicionário ou repositório e poderiam ser constantemente atualizadas, caso exista a necessidade.

- Ações Atômicas: Uma ação simples, que não possui uma "palavra-chave" em sua estrutura.

- Cenário: A história em si. A diferença entre "Normal" e "Excepcional" sendo a própria natureza da narrativa. Uma história inicial seria definida como um "Cenário Normal", já uma variação de seu fluxo de eventos, criando uma nova versão, pode ser classificada como um "Cenário Excepcional".

- Objetivo: O processo de negócio a ser elicitado e modelado.

O modelo de cenários adaptado é descrito na Figura 9.

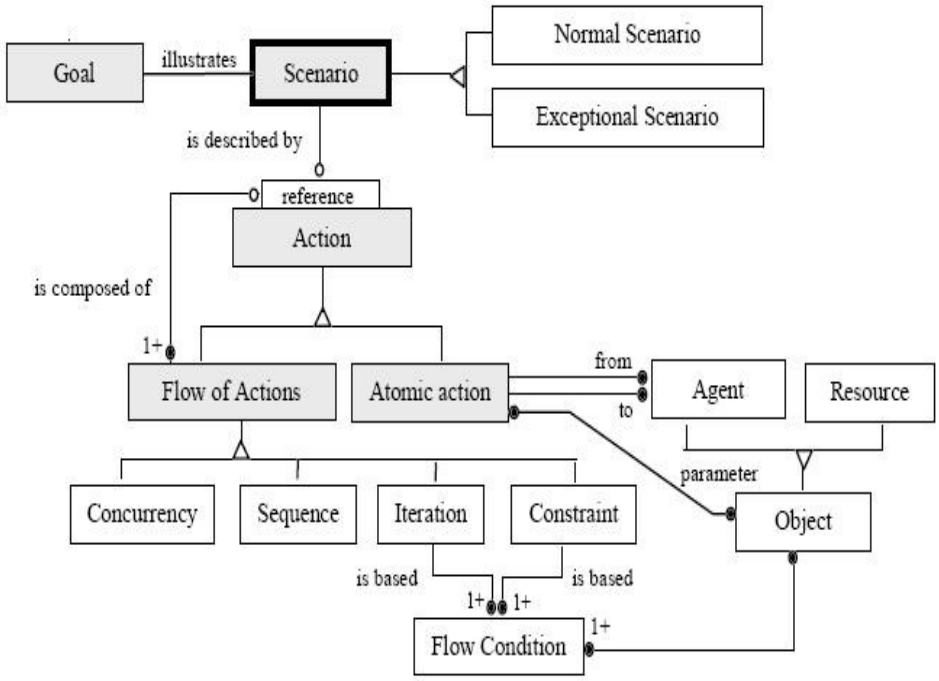


Figura 9 – Modelo CREWS (ACHOUR, 1998) adaptado

Com este objetivo em vista, técnicas de Mineração de Texto (FELDMAN & SANGER, 2007) são utilizadas para a identificação dos elementos básicos de processos presentes na narrativa e suas relações lógico-temporais. A aplicação dos algoritmos necessários para este fase se divide em quatro etapas, descritas a seguir:

5.4.2.1. Tokenização

Nesta etapa, os componentes básicos do texto das histórias são extraídos, compostos pelos parágrafos e frases dos eventos das histórias. O texto é processado e separado em palavras, além de elementos indicativos da estrutura textual como pontuação.

Esta fase inicial é importante, já que define o escopo do processo de extração de informações e prepara o texto puro para o processamento computacional, sendo também considerada como crítica para a eficiência de todo o método de aquisição de elementos relevantes para o processo de negócio.

5.4.2.2. Análise Léxica e Morfológica

A lista separada de palavras, pontuação e outros elementos de texto são agora analisados, de modo a aprofundar o processo de estruturação do texto. O processo pode incluir a análise dos radicais das palavras, detecção de palavras com letras maiúsculas e minúsculas e a classificação gramatical de cada palavra extraída.

O objetivo nesta fase é ampliar a estrutura básica de palavras para uma classificação morfológica de cada palavra do texto, adicionando informação gramatical a cada elemento da lista, para utilização nas fases posteriores.

5.4.2.3. Análise Sintática

A transformação do texto em uma estrutura passível de processamento, iniciada na fase anterior com a classificação gramatical de palavras, é concluída nesta etapa. Com base na classificação morfológica de cada palavra, grupos delas são classificados em termos de Sintagmas Nominais e Sintagmas Verbais, além de Frases (compostas pelos dois tipos de Sintagmas) e outros tipos de estruturas mais complexas. Esta classificação mais refinada transforma o texto em uma estrutura de diversos níveis, onde cada elemento possui informação morfo-sintática suficiente para a identificação de padrões.

A Figura 10 apresenta um exemplo de texto produzido após a análise sintática.

```
(CS
 (S
  (NP O/ART gerente/N de/PREP|+ projeto/N)
  (VP prepara/V juntamente/ADV)
  (NP com/PREP)
  (NP a/ART equipe/N))
 (CNJS
  do/KS
  (NP Escritório de Processos/N)
  (NP os/ART documentos/N para/PREP)
  (NP a/ART reunião/N)))
('.', '.')
```

Figura 10 – Exemplo de trecho de texto após a fase de análise sintática

5.4.2.4. Análise de Domínio

Nesta última etapa da fase de extração automática de conhecimento, a estrutura criada pelas etapas anteriores será processada em busca de padrões gramaticais que possam se adequar ao *template* utilizado como guia para a descoberta de candidatos a elementos de processos. Por exemplo, Sintagmas Nominais são candidatos a Agentes do modelo CREWS e Sintagmas Verbais, candidatos a ações. A extração de elementos do *template*,

de seus relacionamentos e o posterior preenchimento do *template* de cenário (i.e. o processo de negócio em si), conforme vistos nas operações típicas de EI encerram a segunda fase do método, tendo como produto final (vide Figura 11) candidatos a elementos de processo de negócio de forma estruturada.

<ATIVIDADE>
<ATOR>O gerente de projeto </ATOR>
<ACAO>envia </ACAO>
<PARAMETRO>email para o cliente para agendamento da reunião de kick-off </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>O cliente </ATOR>
<ACAO>expõe </ACAO>
<PARAMETRO>a sua necessidade </PARAMETRO>
</ATIVIDADE>

Figura 11 – Exemplo de arquivo de log do processo de extração

5.4.3. Terceira Fase: Construção de representação formal

Finalmente, na terceira e última fase do método, os candidatos a elementos extraídos do *template* de cenário serão convertidos para uma notação de processos de negócio, selecionada para fácil visualização e compreensão pelo analista e pelos participantes do método. Este proto-modelo de processo de negócio será utilizado pelo analista para a criação do modelo final, sendo representativo do conhecimento presente sobre o processo a ser modelado, na narrativa colaborativa criada pelos participantes na primeira fase do método. Devido à possibilidade de modelos alternativos serem utilizados, já que o proto-modelo pode indicar elementos pertencentes aos diferentes pontos de vista sobre o processo, o modelador pode apresentar os modelos alternativos aos participantes e promover a discussão aberta sobre sua validade e opções possíveis para a criação de um modelo definitivo e final.

Para este fim, o modelador utiliza sua experiência na área para guiar o debate e a história original, com seus comentários e informações complementares podem ser

utilizadas como referência para justificar pontos de vista e opiniões. No final desta última etapa, um modelo final de qualidade será criado, com base no consenso dos participantes.

No próximo capítulo, as considerações técnicas para a implementação e suporte ao método Story Mining serão detalhadas, assim como as técnicas, algoritmos e softwares utilizados.

Capítulo 6 – Ambiente de Suporte ao Método StoryMining

6.1. Introdução

No capítulo anterior, o método Story Mining foi definido em detalhes, e a discussão teórica de cada uma de suas três fases foi realizada. Neste capítulo, a discussão se dará sobre a parte prática do método e principalmente, as ferramentas e algoritmos utilizados para apoiá-lo e automatizá-lo.

A implementação do método foi realizada utilizando três softwares: A ferramenta ProcessTeller, para narrativas colaborativas orientada a processos de negócio, O Sistema de Gerenciamento de Workflows Científicos VisTrails, para as rotinas de Mineração de Texto e a ferramenta BizAgi para visualização de processos de negócio descrito em formato de arquivo XPDL, em notação BPMN (BPMN, 2010). A Figura 12 mostra o método e as aplicações utilizadas em cada fase.

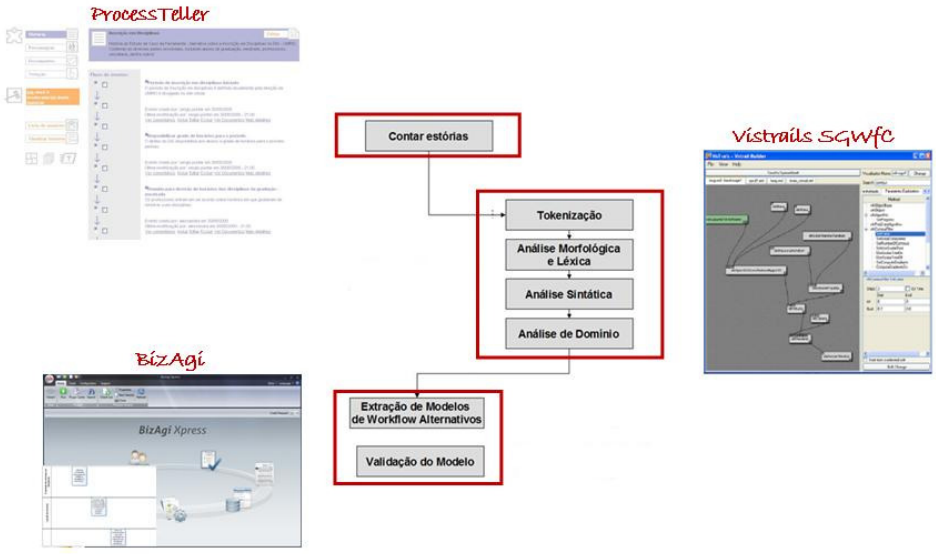


Figura 12 – Método StoryMining com as aplicações utilizadas

6.2. A Ferramenta ProcessTeller

O TellStory (LEAL, BORGES, SANTORO, 2004) foi desenvolvido como uma ferramenta web para suporte à técnica de Group Storytelling. Suas principais funcionalidades são: suporte à criação de histórias, cadastro de usuários, inclusão de eventos, modificações do fluxo de eventos de uma história, além do cadastro de personagens, documentos relacionados à narrativa e a criação de votações.

Além destas funcionalidades básicas, o software ainda permite a criação de uma segunda versão de uma mesma narrativa, além de uma interface para a edição de um texto final a partir do seu fluxo de eventos e a indicação de conhecimento tácito no mesmo texto.

Uma versão da ferramenta foi utilizada, com algumas modificações implementadas especificamente para garantir o suporte ao processo de mineração e, ao mesmo tempo, preservar a expressão livre de conhecimento do Group Storytelling. A nova ferramenta, denominada ProcessTeller, foi adaptada especificamente para uso no Story Mining, se

tornando um ferramenta de Group Storytelling voltada a processos de negócio. A Figura 13 mostra a tela principal de uma história da ferramenta ProcessTeller.

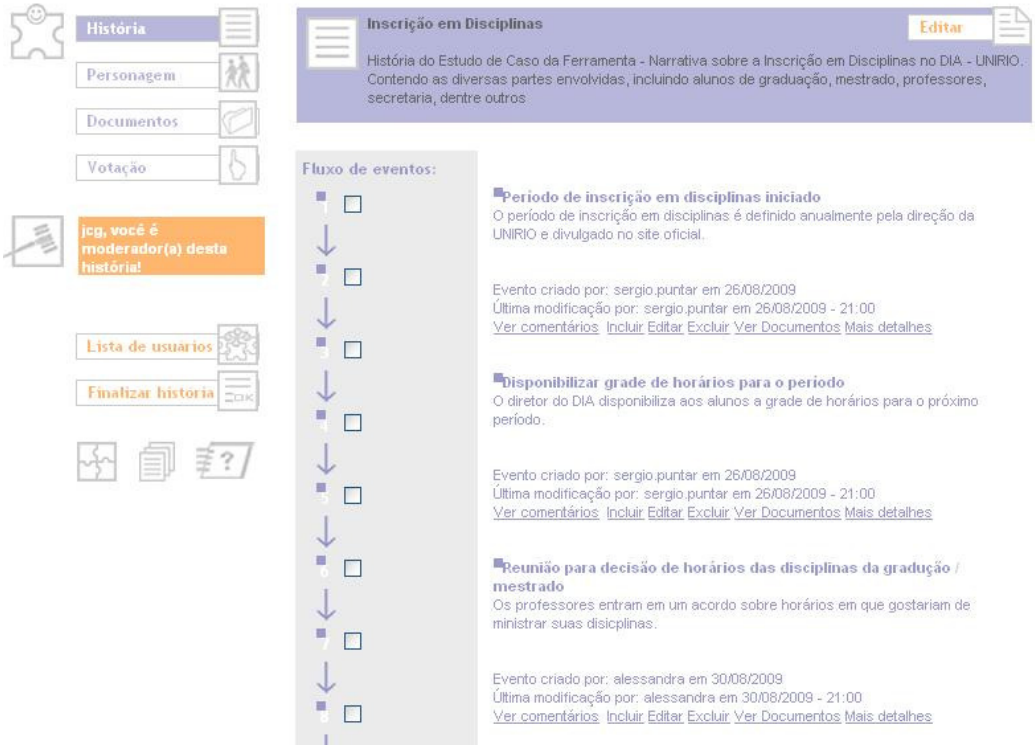


Figura 13 - A ferramenta ProcessTeller

Novas funcionalidades foram implementadas de forma a dar suporte ao método proposto. Estas se encontram descritas a seguir.

6.2.1. Classificação das narrativas utilizando Grupos de histórias

Originalmente, a ferramenta TellStory permite aos usuários adicionar histórias diferentes ao mesmo conjunto de narrativas. Fazendo isto, histórias de diferentes assuntos (descrições de pagamento, contos de fadas, relatórios de usuários, dentre outros) são manipuladas como se fossem parte do mesmo conjunto, prejudicando as tentativas de processamento automáticos do repositório de histórias, aumentando problemas comuns da mineração de texto como, por exemplo, o chamado "ruído" da extração. Por exemplo, se o método for executado para a modelagem de um processo de pagamento, elementos de assuntos que não tem muito em comum com o objetivo, como por exemplo, "contos de fadas" estariam no mesmo conjunto de histórias.

Portanto, uma estrutura de diretório de histórias, organizada por tema, foi implementada, como tentativa de diminuir o impacto deste problema (Figura 14).

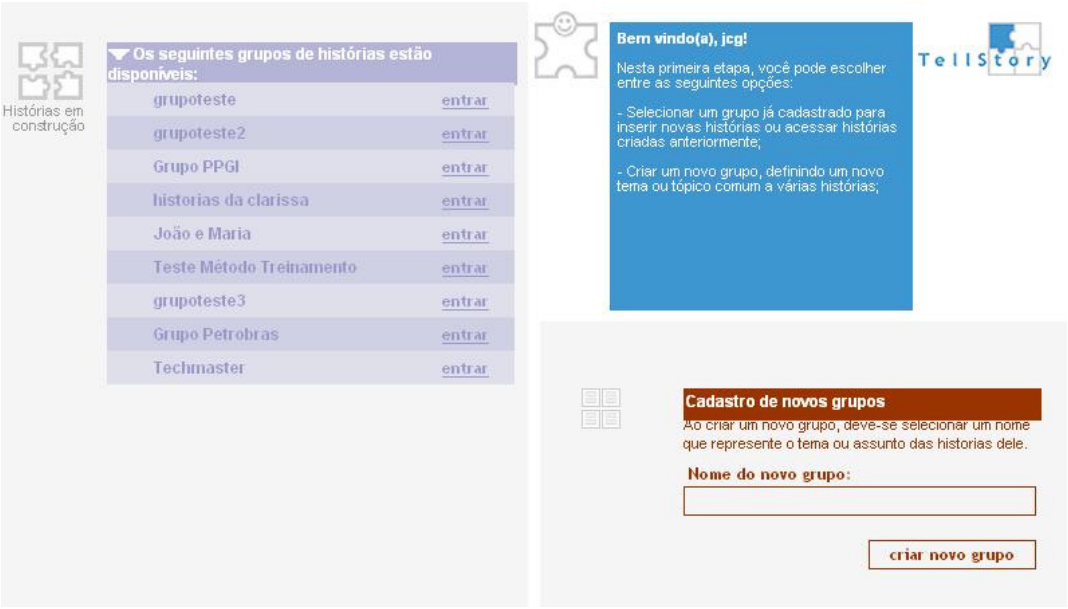


Figura 14: Interface principal do TellStory, com os grupos de histórias

6.2.2. Vinculação de documentos aos eventos da história

A seleção de trechos de histórias extraídas do repositório na primeira fase do método é realizada com base nos "termos de domínio", representativos da linguagem do processo a ser modelado, extraídos dos documentos relacionados às narrativas e presente no conteúdo dos eventos das histórias do repositório. Este é um componente importante da abordagem, estando diretamente relacionado à sua eficiência.

Na ferramenta original, existia apenas a possibilidade de inserir documentos relacionados à história inteira. Em ProcessTeller, a funcionalidade original foi preservada, mas também se possibilita a vinculação dos documentos previamente adicionados à história aos eventos da mesma história (Figura 15).

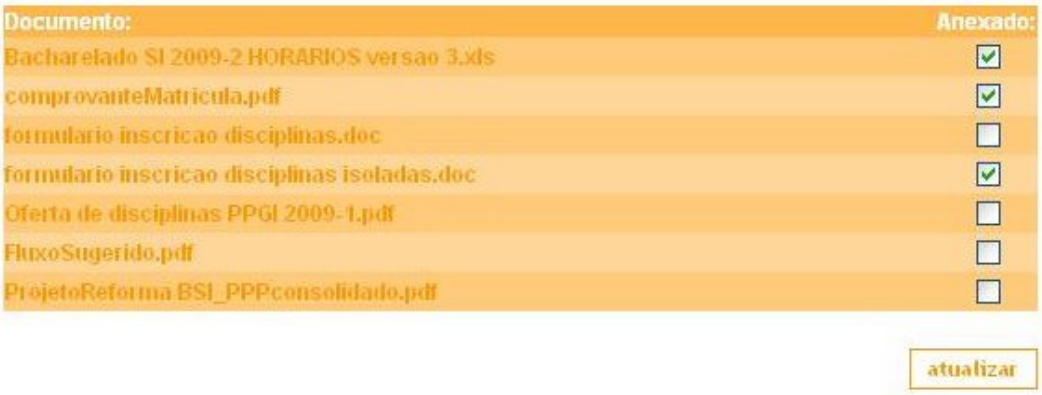


Figura 15: Lista de documentos vinculados a um evento específico

6.2.3. Detecção de Personagens ao nível de eventos da história

A identificação dos atores do processo é uma parte importante das tarefas de elicitação e modelagem de processos, já que elas indicam quais atividades devem ser representadas no modelo final. De forma semelhante, o método Story Mining usa essa informação para enfocar partes da história, geralmente uma frase ou um parágrafo, que pode fornecer conhecimento relevante.

De modo a não interferir na característica livre da narrativa da técnica Group Storytelling, foi escolhida a detecção de personagem dentro de um evento da história como funcionalidade adicional, em detrimento de alternativas, tais como, forçar o usuário a incluir personagens, por exemplo. No entanto, se a ferramenta detectar que um evento não contém nenhum dos personagens previamente cadastrados na história, consulta o usuário se ele desejaria registrar um novo personagem. Desta forma, o usuário é apenas notificado, ao invés de forçá-lo a fazer algo. A notificação pode ser vista na Figura 16.

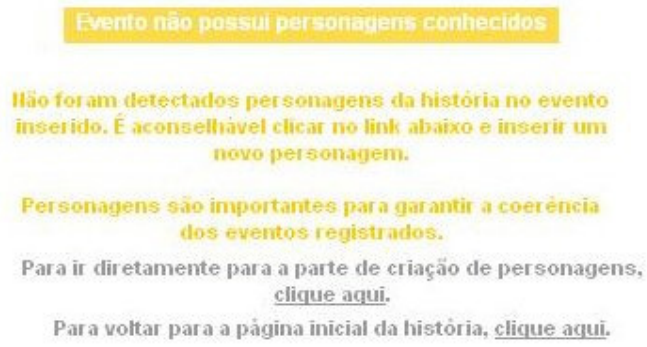


Figura 16: Tela de aviso para um evento que não contém personagens registrados

6.2.4. Evento inicial da história

Devido à natureza orientada a objetivo de um modelo de processo de negócio, um novo campo da história, denominado "Evento Inicial", foi implementado no momento da criação de uma nova história. O usuário pode descrever fatos que permitiriam os eventos da história ocorrer, além de outras informações sobre o contexto da narrativa. Mais uma vez, o objetivo desta nova funcionalidade foi apenas uma maior aproximação do foco da ferramenta ao seu uso na elicitação de processos de negócio, também tomando cuidado em não interferir com o processo de narrativa livre.

6.3. O Sistema de Workflows Científicos VisTrails aplicado à MT

O processo de mineração de textos pode ser visto como um workflow científico (OLIVEIRA et al., 2008). Devido à grande variedade de técnicas, algoritmos e programas de Mineração de Texto de Processamento da Linguagem Natural a serem utilizados em cada passo do método proposto, um sistema de gerenciamento de workflows científicos (SGWfC) foi utilizado para apoiar a definição e execução do workflow . Um workflow científico pode ser definido como a combinação de dados e seqüência de programas (CAVALCANTI, TARGINO, BAIÃO *et al*., 2005). Uma vez executados, a sequência de programas e suas entradas e saídas caracterizam um experimento.

Um SGWfMC permite a execução e a análise comparativa entre diversas instâncias do workflow, no qual pode-se varias as técnicas, programas, valores de argumentos de cada algoritmo e até mesmo fontes de conhecimento e entradas para o método como um todo.

Para a implementação das técnicas de Mineração de Texto, foi adotado o SGWfC VisTrails (CALLAHAN, FREIRE E SANTOS, 2006), devido às suas funcionalidades avançadas para visualização e comparação de resultado, além de sua extensibilidade nativa para chamar funções e aplicações externas, por meio de módulos extras, que podem ser desenvolvidos como classes em linguagem Python (LUTZ, 2006) e inseridos no VisTrails como bibliotecas extras de módulos para workflows científicos. Para as técnicas de Mineração de Texto, foram utilizados o framework NLTk (BIRD & LOPER, 2004) e a plataforma Biguá (OLIVEIRA, 2008). O Biguá é uma plataforma de desenvolvimento de aplicações de mineração de texto, com suporte aos idiomas português e inglês. Possui duas versões de execução, uma *stand-alone* e a outra utilizando um formato de workflow, onde as partes são definidas e suas execuções coordenadas para aplicação das técnicas de mineração de texto.

A versão "Workflow Biguá" (OLIVEIRA, 2008) foi utilizada para a implementação do pré-processamento e da aplicação de técnicas de cálculo de freqüência de palavras nos documentos relacionados às histórias. Atualmente, a plataforma possui o nome de MiningFlow, inclusive estando disponível no site Sourceforge. O framework NLTk – Natural Language Toolkit (BIRD & LOPER, 2004) foi utilizado devido à vasta gama de algoritmos de Parsing Sintático, Shallow Parsing e Tokenização, além de possuir diversos corpus de documentos etiquetados em múltiplos idiomas.

Durante o trabalho de pesquisa desta dissertação, foram desenvolvidos três módulos do VisTrails. Os dois primeiros, são "frontends" para utilização dos métodos e classes das bibliotecas Workflow Biguá e NLTk com o SGWfC. O último é a biblioteca do método StoryMining, onde funções complementares, como conversão entre diversos formatos de texto, filtragem por listas de palavras, dentre outras funcionalidades, estão presentes para uso.

A utilização conjunta das bibliotecas com um sistema de workflow científico possibilita um reuso de código e, mais ainda, a reprodução do método Story Mining em outras áreas e condições, além de uma maior facilidade em futuras pesquisas e aplicações nesta linha. A Figura 17 ilustra o workflow definido, com base na interface do VisTrails.

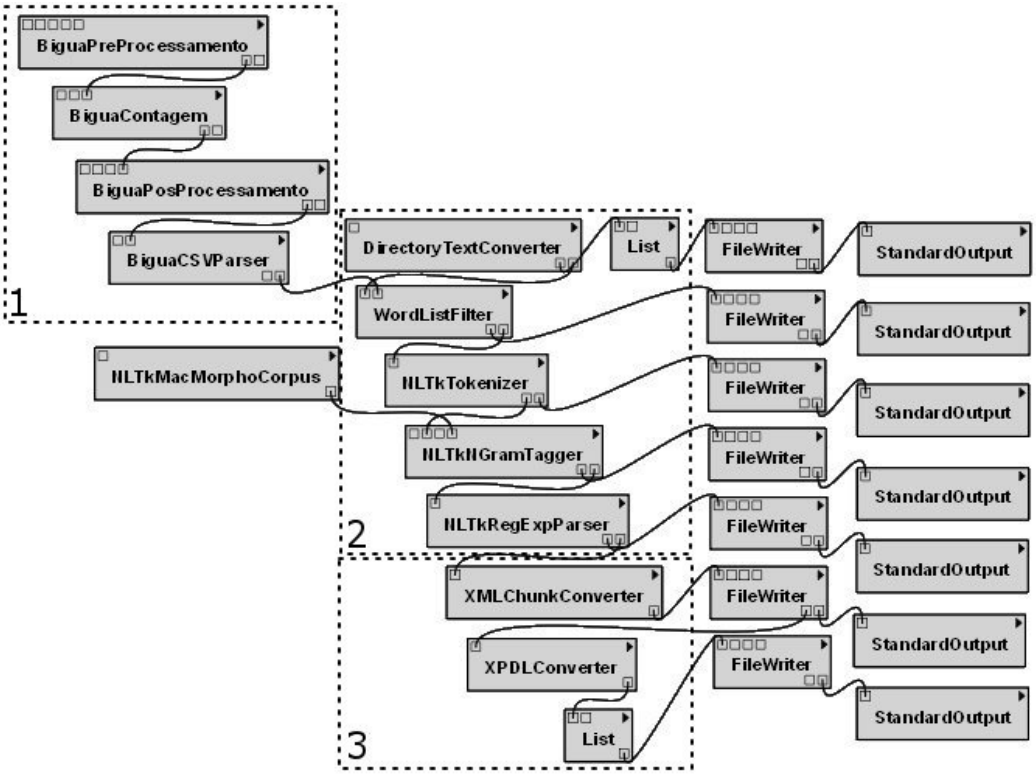


Figura 17 – Workflow utilizado para o método

6.3.1. Descrição do Workflow do Método

O Workflow utilizado na implementação do método, mostrado na Figura 17, possui três partes distintas, marcadas também na figura 16 por quadrados tracejados. Na primeira, as técnicas de cálculo de freqüência de palavras e radicalização de termos é aplicada nos documentos relacionados à história, inseridos pela interface do ProcessTeller. A segunda parte retira o texto dos eventos da história, converte para texto puro e aplica uma seqüência de técnicas de pré-processamento, desde a filtragem de frases relevantes até a análise sintática e a estruturação final do texto. A terceira e última parte é a conversão do texto estruturado em formatos específicos para processos de negócio.

As partes referentes aos módulos FileWriter e StandardOutput, vinculadas a cada classe da servem para a geração de arquivos de log parciais do processo como um todo ou dos dois arquivos finais de saída do método. Estes logs são importantes para a análise de possíveis erros e desempenho baixo do método e para avaliar técnicas utilizadas, mas não são considerados como parte do processo em si, sendo apenas módulos básicos para escritas de arquivos em disco.

6.3.1.1. Primeira parte: Processamento dos documentos das histórias

Nesta fase inicial, o download dos documentos da história é realizado e o local indicado para a ferramenta Workflow Biguá. Três módulos são utilizados (BiguaPreProcessamento, BiguaContagem e BiguáPosProcessamento) para as operação desta etapa.

A extração de palavras é realizada, através do pré-processamento dos documentos e contagem da freqüência de cada termo, utilizando o cálculo TF-IDF (Term Frequency-Inverse Document Frequency) para toda a coleção de documento presente.

De modo a evitar a ambigüidade entre palavras com significado semelhante, uma técnica adicional, chamada de RSLP Stemmer (ORENGO & HUYCK, 2001), realiza a redução das palavras aos seus radicais, refinando a filtragem de palavras de domínio do método. Esta estratégia já foi utilizada em outras disciplinas como Medicina (HOLZINGER, GEIERHOFER, MÖDRITSCHER *et al.*, 2008) e com outras aplicações como a geração de Topic Maps a partir de texto (BIEMANN, QUASTHOFF, BÖHM *et al.*., 2003). Alternativamente, outras formas mais complexas de conhecimento externo poderiam ser utilizadas para a filtragem, como ontologias, tesauros e dicionários, dentre outras.

Esta operação gera um arquivo de formato CSV, contendo os resultados do processo. O módulo do workflow BiguaCSVParser é utilizado para converter os termos encontrados para uma lista de palavras, sendo este o resultado final desta etapa.

6.3.1.2. Segunda parte: Processamento dos eventos das histórias

A segunda parte se inicia com o módulo DirectoryTextConverter, que converte o conteúdo dos eventos da história a ser processada para um arquivo de texto puro. Um filtragem de frases então é realizada, pelo módulo WordListFilter, com base na lista de palavras encontrada na primeira parte do workflow. Este módulo mantém apenas as frases dos eventos que contém pelo menos uma palavra cujo radical ou forma canônica esteja presente dentre os resultados da primeira parte.

Em seguida, a técnica de tokenização é aplicada (NLTkTokenizer), gerando uma lista de palavras e pontuação a partir do texto filtrado. Segue então a criação da estrutura gramatical, para este fim, inicialmente utiliza-se um Trigram Tagger, com uma tag default de “N” (“Nome” no corpus MAC-MORPHO) para palavras não classificadas (NLTkNGramTagger), em conjunto com o corpus de treinamento MAC-MORPHO

(ALUISIO, PINHEIRO, MANFRIM, 2004), pelo módulo NLTkMacMorphoCorpus, para classificar cada palavra.

Esta classificação inicial é refinada pelo módulo NLTkRegExpParser, que aplica a técnica de *Shallow Parsing* (OSBORNE, 2000) (SANG & BUCHHOLZ, 2000), a partir de uma gramática de expressões regulares. Uma estrutura mais refinada e de diversos níveis como resultado e serve como base para a parte final do workflow.

6.3.1.3. Terceira parte: Extração de Informação

Na última parte, os elementos extraídos do texto estruturado pelas fases anteriores serão processados e os candidatos a elementos de processo serão extraídos e convertidos para duas formas específicas.

Os elementos do processos a serem extraídos, para a implementação da pesquisa desta dissertação foram: Atores, Atividades e os Parâmetros envolvidos. Esta escolha se deu, principalmente por estes serem os principais elementos de um processo de negócio genérico.

O primeiro formato, gerado pelo módulo XMLChunkConverter, é uma estrutura similar a um XML, onde se destaca cada atividade, com seus respectivos ator e parâmetro, além de possíveis palavras que indiquem algum tipo de fluxo de ações, caso existam na frase. Esta primeira saída é um arquivo semelhante a um "log" de sistema, onde o resultado bruto da extração de conhecimento se encontra em sua forma mais pura.

O segundo formato é um arquivo em formato XPDL (XPDL, 2010), gerado pelo módulo XPDLConverter, a partir da primeira saída desta parte do workflow. Devido ao seu formato, ele pode ser carregado em diversas ferramentas de modelagem de processos de negócio. Ele permite que os elementos extraídos sejam visualizados na forma de raias e atividades, tornando a visualização e posterior manipulação destes elementos e a modelagem de um processo final se torne mais fácil e em uma linguagem mais próxima à de que um analista de processo está acostumado.

6.4. Visualização de Proto-Modelo em notação BPMN pelo BizAgi

A visualização dos resultados finais é parte importante de uma aplicação de Mineração de Texto e ela deve se adequar ao usuário final do sistema. No caso da Modelagem de Processos de Negócio, as operações realizadas pelo workflow científico geram um

arquivo de saída passível de ser utilizado com uma ferramenta de modelagem tradicional.

Foi constatado, durante a experimentação de vários tipos de ferramenta para modelagem de processos de negócio, que cada uma das ferramentas utilizava um padrão XPDL customizado, sendo bastante difícil a geração de um arquivo com este formato, sem o vínculo com um software específico.

Para a implementação desta parte do método, foi utilizada a ferramenta BizAgi Process Modeler (BIZAGI, 2009). A escolha se justifica pelo fato de possuir uma versão gratuita com razoável quantidade de funcionalidade e possibilidade de carregar arquivos XPDL. A notação utilizada pela ferramenta é a BPMN – Business Process Modeling Notation (BPMN, 2010).

BPMN provê uma notação gráfica para a especificação de processos de negócio em um Business Process Diagram (BPD). O objetivo da notação é prover uma representação que seja intuitiva para os usuários e, ao mesmo tempo, representa a complexa semântica dos processos. A Figura 18 demonstra uma parte de um arquivo XPDL gerado pelo método e carregado pelo BizAgi.

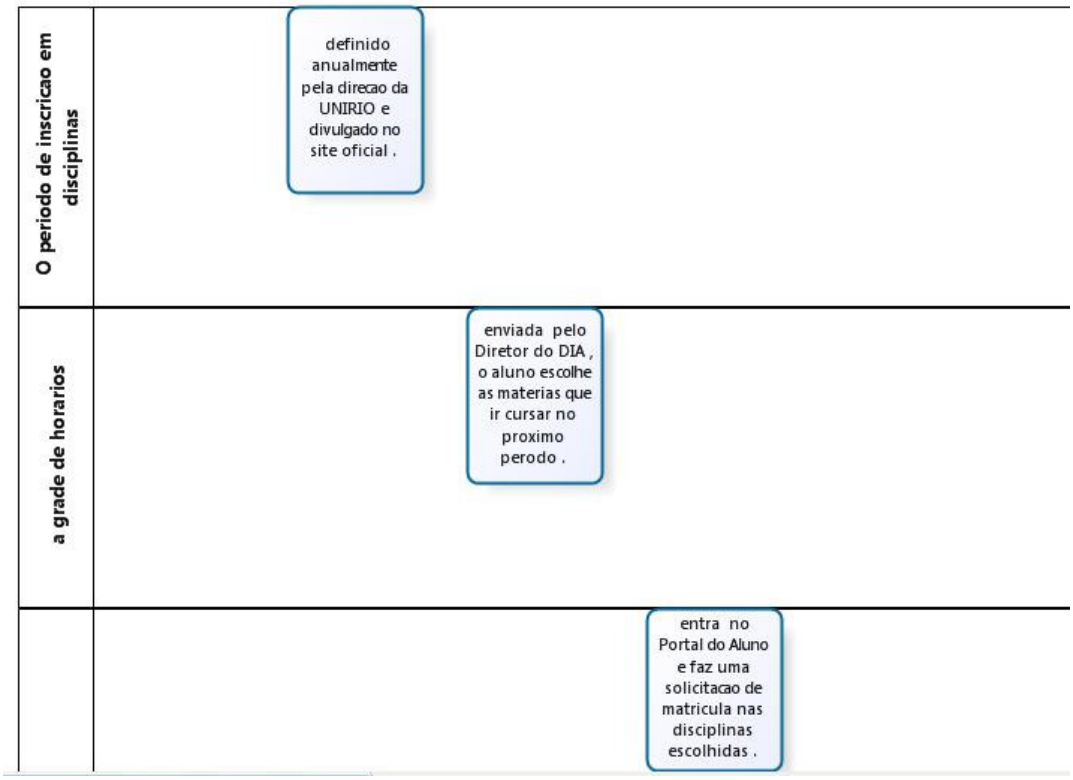


Figura 18 – Visualização do proto-modelo, mostrando atividades e atores candidatos

Com o uso dos softwares descritos, as três fases do método (Construção coletiva de histórias, Mineração de histórias e Criação de representação formal), detalhadas no capítulo 5, podem ser aplicadas na prática. Esta implementação específica foi utilizada para avaliar a viabilidade do método em um cenário real de Elicitação de Modelagem de Processos de Negócios em organização e processos de diferentes perfis, de modo a extrair evidências sobre pontos fortes e fracos da solução e conclusões gerais sobre seu uso.

No próximo capítulo, os estudos de casos da pesquisa e suas avaliações (GONÇALVES, SANTORO, BAIÃO, 2010a) serão descritos, de modo a discutir os resultados da aplicação do método Story Mining.

Capítulo 7 – Estudos de Caso e Análise de Resultados

7.1. Metodologia

A definição de um estudo de caso, de acordo com YIN (2005), "é uma investigação empírica que investiga um fenômeno contemporâneo dentro de seu contexto da vida real, especialmente quando os limites entre o fenômeno e o contexto não estão claramente definidos". O mesmo autor apresenta ainda quatro aplicações possíveis para o método do Estudo de Caso:

1. Para explicar ligações causais nas intervenções na vida real que são muito complexas para serem abordadas pelos Surveys ou pelas estratégias experimentais;
2. Para descrever o contexto da vida real no qual a intervenção ocorreu;
3. Para fazer uma avaliação, ainda que de forma descritiva, da intervenção realizada;
4. Para explorar aquelas situações onde as intervenções avaliadas não possuam resultados claros e específicos.

No caso desta dissertação, o estudo de caso foi selecionado para avaliação da proposta devido à complexidade da situação prática abordada, envolvendo processos de negócio e gestão de conhecimento, com o objetivo de demonstrar como o método Story Mining pode ser considerado viável para a Elicitação de Processos com vantagens nas situações descritas.

Em cada estudo de caso, foi a aplicado o método Story Mining em suas três fases com um grupo específico de narradores. Na seleção dos narradores, objetivo foi selecionar pessoas de todos os grupos que (a) executem tarefas relativas ao processo em questão, (b) tenham contato direto com o processo a ser elicitado, (c) visualizam parte do processo mesmo sem participar deste.

Estes três grupos possibilitam reunir o maior número de pontos de vista possível, o que pode levar a uma maior qualidade dos produtos gerados pelo método como um todo.

O período de tempo para cada fase variou de uma semana a um mês. Principalmente para a primeira fase, que consiste na narrativa pelos participantes, por ser um processo colaborativo, dinâmico e que envolve o fator humano, se torna difícil precisar uma quantidade de tempo para execução do processo de Group Storytelling.

Após a construção das narrativas, o workflow científico foi executado pelo VisTrails foram aplicados, em um processo automático que dura cerca de 2 horas. A revisão destes resultados e o trabalho posterior do analista (modelagem que leva ao produto final) e a interação entre analistas e participantes, para a discussão do resultado e possível consenso, pode levar bastante tempo, sendo difícil precisar um período específico.

O workflow foi aplicado duas vezes para cada estudo de caso. Na primeira vez, foi utilizada a primeira gramática de expressões regulares (Gramática I) para estruturar o texto, visando criar uma estrutura detalhada, inclusive separando palavras que indicam fluxos de ações mais complexas do modelo CREWS.

A segunda aplicação utilizou uma outra gramática (Gramática II), que tem como objetivo extrair frases inteiras de atividades, com um enfoque menor em detalhar a estrutura do texto. Para fins de visualização e modelagem, esta gramática foi selecionada como mais útil para a geração do proto-modelo em formato XPDL.

A avaliação dos resultados de cada estudo de caso foi realizada de duas formas: a primeira, pela avaliação do resultado direto do processo de mineração e da análise das atividades extraídas. Nos casos onde existia um processo pré-modelado que pertencia ao escopo do processo alvo do estudo, foi realizada uma comparação entre as atividades presentes no modelo extraído e no modelo pré-existente.

O conteúdo de cada atividade a ser comparada foi analisado e dois grupos foram criados: atividades coincidentes e atividades não-coincidentes. A natureza das atividades coincidentes também foi avaliada e, no caso das atividades não-coincidentes, uma análise mais detalhada foi realizada para extrair evidências sobre os motivos de sua ausência em um modelo e a presença em outro.

Além disso, foi utilizada a análise dos comentários da ferramenta ProcessTeller e o uso de questionários, para adquirir a opinião dos participantes sobre o método e a ferramenta de suporte e os resultados obtidos (modelo gerado).

Desta forma, foi possível captar as impressões de cada um dos envolvidos, de modo a esclarecer percepções e evidências da análise preliminar das atividades.

7.2. Estudo de Caso 1: Processo de Inscrição em Disciplinas DIA/UNIRIO

7.2.1. Descrição

O primeiro estudo de caso foi planejado para servir como uma etapa de avaliação da solução proposta pela pesquisa, além de extrair evidências sobre o método e prover novas idéias sobre os fatores envolvidos no processo colaborativo. Ele foi realizado no Departamento de Informática Aplicada (DIA) da Universidade Federal do Estado do Rio de Janeiro (UNIRIO).

O processo a ser modelado foi escolhido como "Inscrição em Disciplinas", com todos os seus possíveis nuances conhecidos e desconhecidos (Graduação, Mestrado, etc.). Uma das razões para tal escolha foi sua característica de ser um processo bastante comum em instituições educacionais e conhecido por uma vasta gama de "narradores candidatos", incluindo alunos, professores e funcionários da universidade envolvidos com as suas atividades.

Embora existam diversos tipos de processos de inscrição em disciplinas na instituição escolhida, devido à presença de cursos de graduação e pós-graduação no mesmo centro, uma decisão foi tomada para não se dividir os processos em duas narrativas diferentes, mas, pelo contrário, permitir aos participantes de ambos os contextos contarem seus pontos de vista e experiências em uma mesma história, de modo a avaliar a riqueza de conhecimento e as diferenças presentes na narrativa.

Seguindo o método, os narradores foram selecionados dentre funcionários da universidade, corpo docente, e corpo discente, incluindo alunos de graduação e de mestrado. Esta seleção foi realizada tanto por meio de convites diretos por email feitos aos voluntários destes grupos de pessoas, até uma chamada geral para participação nas listas de email dos cursos deste centro, nos seus respectivos níveis (Colegiado, Alunos, dentre outros).

Os convites diretos se destinaram a dezoito pessoas, incluindo todos os diferentes tipos de narradores envolvidos. Os convites abertos e chamadas gerais para participação se

destinaram a inserir pessoas dos grupos que não haviam sido selecionadas, de modo a ampliar o número de pontos de vista diferentes e de narradores envolvidos com a tarefa.

O grupo selecionado utilizou a ferramenta colaborativa por cerca de um mês, expondo seus pontos de vista sobre o processo de inscrição em disciplinas, incluindo comentários em partes da história inseridas por outros participantes, enviando documentos relacionados à história para o sistema, além de ler toda a narrativa criada colaborativamente, buscando "insights" para suas próprias contribuições.

Na primeira semana de estudo de caso, o primeiro participante a contribuir com a história, um estudante de graduação, criou oito eventos de uma única vez, descrevendo toda sua visão do processo. Uma análise posterior dos resultados do estudo de caso demonstrou que estes eventos iniciais funcionaram como uma "visão essencial" dos componentes básicos do processo. Após a contribuição deste usuário, os outros participantes começaram a comentar seus eventos e inserir os seus próprios, complementando a história com seu conhecimento, geralmente inserindo novos fatos no meio da cadeia de eventos inicialmente descrita, a citada "visão essencial".

Conforme o estudo de caso progrediu após a inserção inicial dos oito eventos, dezoito eventos adicionais foram incluídos ao fluxo principal da história, e um total de 51 novos comentários foram registrados pelos usuários. Conforme a narrativa aumentou em termos de tamanho e riqueza de detalhes, alguns usuários escolheram por conta própria contribuir apenas com comentários em eventos previamente registrados, deixando suas opiniões e "insights" sobre as contribuições dos outros participantes.

Após a segunda semana do estudo, o número de novos eventos reduziu gradualmente e o ritmo de criação de novos comentários começou a aumentar consideravelmente.

Ao fim da primeira fase do método, o fluxo principal da história narrado pelos participantes e a informação auxiliar registrada (personagens, documentos, dentre outros) foi utilizada para a segunda fase, a aplicação de algoritmo de Mineração de Textos e Processamento da Linguagem Natural.

Para esta fase, a fonte principal de mineração foi selecionada: os eventos da história. Os documentos relevantes inseridos pelos usuários foram também escolhidos como entrada principal para aplicação de técnicas de Mineração de Texto para formação de um critério para selecionar frases e textos relevantes, conforme descrito no método. Após a extração e geração do proto-modelo de processo a partir das atividades extraídas, um

questionário foi enviado para os participantes de modo a coletar suas percepções a respeito do método, da ferramenta e do processo de elicitação como um todo.

7.2.2. Análise

O proto-modelo gerado e o log de atividades descobertas foi extraído para ser comparado com um modelo do mesmo processo de negócio (Inscrição em Disciplinas), construído utilizando a técnica tradicional de entrevistas e modelagem realizada pelo analista. Devido à divisão de "trigger words", especificada no template do modelo CREWS (Capítulo 5), foi decidido que uma pré-análise seria realizada, de modo a considerar as atividades em todo o seu contexto, unindo o texto das atividades relacionadas por "trigger words" extraídas pelo método. A Tabela 1 mostra uma comparação geral sobre os dois modelos, sendo PM1 o modelo criado manualmente pelo método tradicional e PM2 o proto-modelo gerado pelo método Story Mining.

O método Story Mining conseguiu descobrir automaticamente 8 das 21 atividades do modelo de processo criado manualmente. Foi observado que 7 das 13 atividades restantes não descobertas de PM1 representam atividades que não são percebidas por usuários (por exemplo, atividades envolvendo a interação com sistemas de informação) e tarefas administrativas (Ex: "Arquivar Relatório de Inscrição de Disciplinas"). A Tabela 1 ilustra estes resultados.

Este fato indica que a diferença principal entre as atividades do modelo gerado pelo método Story Mining e as atividades presentes no modelo pré-existente tem a ver com as percepções das pessoas envolvidas. Por exemplo, atividades relacionadas com o uso de sistemas de informação ou com atividades meramente administrativas foram descritas com sucesso pelo modelo tradicional, enquanto atividades claramente visualizadas pelos participantes do estudo estavam presentes no modelo gerado pelo método.

Tabela 1 – Resultado comparativo entre os dois modelos (Gramática I)

Estatística	PM1	PM2
Total de atividades	21	39
# de atividades coincidentes	8	11
# de atividades não-coincidentes	13	28

O proto-modelo extraído (PM2) também contém 28 atividades que não foram detalhadas ou não estão presentes no processo modelado manualmente (PM1), revelando uma riqueza de atividades previamente desconhecidas, que representavam experiências reais vividas pelos participantes (Ex: Alunos especiais que podem se inscrever em disciplinas isoladas, sem serem membros efetivos do curso de mestrado). Dos elementos de PM2, nove foram extraídos como atividades pelo método, mas que em PM1 estavam representados como regras de negócio (Ex: ”O período de inscrição de disciplinas é definido anualmente”). Esta análise revelou que, embora personagens estejam presentes em uma história, elementos especiais do processo (sistemas de informação, regras de negócios, dentre outros) podem ser descritos de formas diferentes, separados do fluxo principal de eventos, estimulando os narradores a descrevê-los.

Uma segunda comparação foi realizada, com base na classificação do conteúdo de cada atividade extraída. Para esta análise, as atividades foram agrupadas em 4 categorias: (i) Atividades Gerais (Atividades comuns a um processo de inscrição de disciplinas genérico), (ii) Atividades de Mestrado (Atividades específicas pertencentes à inscrição em disciplinas do curso de mestrado), (iii) Atividades de Graduação (Atividades específicas pertencentes à inscrição de disciplinas do curso de graduação) e (iv) Atividades de Aluno Especial (Atividades pertencentes à inscrição de disciplinas dos alunos externos à Universidade).

A Tabela 2 mostra os resultados desta segunda comparação de resultados.

Tabela 2 – Atividades Extraídas por grupo (Gramática I)

Grupo	**# de Atividades (PM1)**	**# de Atividades (PM2)**
Gerais	10	16
Graduação	10	11
Mestrado	1	5
Aluno Especial	0	7
Total de # Atividades	**21**	**39**

Em um nível mais detalhado, uma comparação de cada atividade extraída pelo Story Mining (PM2) e com cada atividade referente ao processo de Inscrição de Disciplinas

do processo validado pré-existente na organização (PM1) foi realizada. Pode-se perceber que diversos elementos de processos (não se restringindo apenas a atividades) que não estavam presentes no modelo validado foram encontrados. A Tabela 3 descreve as atividades do modelo PM1 que são referentes ao processo de "Inscrição em Disciplinas". A Tabela 4 indicam as atividades do modelo PM2 que são relativas as atividades de PM1 e a Tabela 5 mostra quais atividades de PM2 que são diferentes do processo já existente.

Tabela 3 – Atividades do Processo validado (Inscrição em Disciplinas)

Processo Validado	
Item	Descrição do Item
5.1.1.1.1	Acessar portal do aluno
5.1.1.1.2	Solicitar inscrição em disciplinas veterano
5.1.1.1.3	Criticar solicitação de inscrição em disciplinas
5.1.1.1.4	Informar disponibilidade do relatório de solicitação de inscrição em disciplina
5.1.1.1.5	Inscrever em disciplina fora do prazo
5.1.1.1.6	Solicitar impressão da documentação para inscrição em disciplina
5.1.1.1.7	Emitir relatório de solicitação de inscrição em disciplina e histórico escolar
5.1.1.1.8	Agendar semana de confirmação de inscrição em disciplina
5.1.1.1.9	Verificar relatório de solicitação de inscrição em disciplina
5.1.1.1.10	Realizar modificações na solicitação
5.1.1.1.11	Encaminhar relatório de solicitação de inscrição em disciplina
5.1.1.1.12	Verificar modificações na solicitação de inscrição em disciplina
5.1.1.1.13	Registrar modificações na inscrição em disciplinas
5.1.1.1.14	Confirmar inscrição em disciplinas
5.1.1.1.15	Arquivar relatório de inscrição em disciplinas
5.1.2.1.1	Confeccionar proposta de grade horária
5.1.2.1.2	Propor grade horária
5.1.2.1.3	Pronunciar matérias ofertadas da Pós Graduação
5.1.2.1.4	Solicitar mudanças
5.1.2.1.5	Efetuar mudanças necessárias
5.1.2.1.6	Produzir minuta

Tabela 4 – Atividades extraídas relacionadas com o modelo validado (Gramática I)

Atividade Extraída	Atividade do modelo validado
O período de inscrição em disciplinas é definido anualmente pela direção da UNIRIO e divulgado no site oficial	5.1.1.1.3
A grade de horários enviada pelo diretor	5.1.1.1.3
O aluno escolhe as matérias que irá cursar no próximo período	5.1.1.1.3
O aluno entra e faz uma solicitação de matrículas nas disciplinas	5.1.1.1.3
A secretária imprime e entrega as solicitações de matrícula dos alunos aos professores tutores das respectivas turmas	5.1.1.1.3
Os professores informam aos alunos das que são tutores os dias para a confirmação de matrícula as mudanças enviadas pelos professores e confirma a inscrição dos alunos	5.1.1.1.9, 5.1.1.1.10, 5.1.1.1.14
Os professores entram em acordo sobre horários que gostariam de ministrar suas disciplinas	5.1.2.1.4
A diretora organiza uma grade de horários	5.1.2.1.1, 5.1.2.1.2
O horário de disciplinas é divulgado no site	5.1.2.1.3
Programa Há um período de discussão de disciplinas a serem ofertadas	5.1.2.1.4
e tutor se reúnem para verificar o histórico e sugerir alterações em sua inscrição, visando o melhor aproveitamento	5.1.1.1.10

Tabela 5 – Atividades Novas extraídas pelo Método (Gramática I)

Atividade Extraída	Atividade do modelo validado
A diretoria encaminha a mesma para a secretária	Nova
do PPGI O aluno de mestrado tem que entregar o documento de inscrição de disciplinas	Mestrado
O orientador assina o termo de compromisso de orientação para validar sua orientação junto com o aluno de mestrado	Mestrado
Se o aluno não confirmar a inscrição em disciplinas com o professor tutor, será considerado abandono de curso por parte	Nova
O aluno especial pode escolher cursar disciplinas isoladas antes de ingressar no curso	Aluno Especial
Podem ser cursadas até 2 disciplinas por semestre	Nova
Não são todas as disciplinas que podem ser cursadas por alunos especiais	Aluno Especial
O aluno especial deve anexar à documentação comum para inscrição de disciplinas isoladas uma carta explicando o motivo pelo qual deseja cursar a disciplinas	Aluno Especial
O aluno especial deve comparecer na secretaria para confirmar se seu pedido foi DEFERIDO	Aluno Especial
definem as disciplinas que serão cursadas no semestre	Nova
O aluno de mestrado tem que escolher as disciplinas juntamento com o seu orientador	Mestrado
do aluno é realizada no sistema após a entrega	Nova
assinado pelo orientador	Nova
O aluno ainda tem um período para inclusão	Nova

do PPGI pode ser feita num prazo de 30 dias após o fim	Mestrado
O aluno que trancar uma disciplina até esse período não é considerado reprovado	Nova
Se não trancar deve cursar até o fim ou será considerado abandono	Nova
Há um período de discussão de disciplinas a serem ofertadas	Nova
uma grade preliminar pela Escola	Nova
O professor responsável pela disciplinas deve avaliar o pedido de matrícula em disciplina isolada e encaminhar notificação de deferimento	Aluno Especial
O professor responsável pela disciplina tem total autonomia para deferir	Aluno Especial
as inscrições em disciplina isolada deferidas, para posterior emissão de declaração	Aluno Especial
os alunos para confirmar a matrícula em disciplinas	Nova
e encaminha as mudanças à secretaria	Nova
Alunos de mestrado precisam da assinatura	Mestrado
O aluno também pode, durante período prédeterminado pela UNIRIO PROGRAD) trancar a matrícula do curso	Nova
do o trancamento é de	Nova
e não pode ser feito no primeiro semestre	Nova

Após a extração e comparação de resultados, a segunda gramática foi utilizada para esclarecer as atividades extraídas, novamente aplicada no texto das histórias, iniciando o mesmo processo de mineração do workflow científico implementado. O resultado está exposto na Tabela 6.

Tabela 6 – Atividades extraídas com o uso da Gramática II

Atividade Extraída
O período em inscrição em disciplinas é definido anualmente pela direção da UNIRIO e divulgado no site oficial.
A grade horários enviada pelo Diretor do DIA, o aluno escolhe as matérias que irá cursar no próximo período.
O aluno entra no Portal do Aluno e faz uma solicitação de matrícula nas disciplinas escolhidas.
A secretária imprime e entrega as solicitações de matrícula dos alunos aos professores tutores das respectivas turmas.
Os professores informam aos alunos das turmas que são tutores os dias para a confirmação de matrícula.
As mudanças enviadas pelos professores e confirma a inscrição dos alunos.
Os professores entram em um acordo sobre horários em que gostariam de ministrar suas disciplinas.
A diretora organiza uma grade de horários.
A diretora encaminha a mesmo para a secretária
O horário das disciplinas é divulgado no site do PPGI O aluno de mestrado tem que entregar o documento de inscrição em disciplinas e o termo de compromisso de orientação.
O orientador assina o termo de compromisso de orientação para validar sua orientação junto com o aluno de mestrado.
O aluno não confirmar a inscrição em disciplinas com o professor tutor, será considerado abandono de curso por parte do aluno.
O aluno especial pode escolher cursar disciplinas isoladas antes de ingressar no curso.
Podem ser cursadas ate 2 disciplinas por semestre, por no máximo 2 semestres

Não são todas as disciplinas que podem ser cursadas por alunos especiais.
O aluno especial deve anexar à documentação comum para inscrição em disciplinas isoladas uma carta explicando o motivo pelo qual deseja cursar a disciplina.
O aluno especial deve comparecer na secretaria para confirmar se seu pedido foi DEFERIDO ou INDEFERIDO.
Definem as disciplinas que serão cursadas no semestre
O aluno de mestrado tem que escolher as disciplinas juntamente com seu orientador.
A inscrição de disciplinas do aluno é realizada no sistema após a entrega do requerimento de inscrição em disciplinas na secretaria , assinado pelo orientador e pelo próprio aluno .
O aluno ainda tem um período para inclusão ou exclusão de disciplinas.
Exclusão de disciplinas no PPGI pode ser feita num prazo de até 30 dias após o fim do período regular de matrícula.
O aluno que trancar uma disciplina até este período não é considerado reprovado.
Se não trancar deve cursar até o fim, ou será considerado abandono.
Há um período de discussão de disciplinas a serem ofertadas e horários, com base em uma grade preliminar enviada pela Escola /
Programa . Há um período de discussão de disciplinas a serem ofertadas e horários , com base em uma grade preliminar enviada pela Escola /
O professor responsável pela disciplina deve avaliar o pedido de matrícula em disciplina isolada e encaminhar notificação de deferimento ou não do pedido .
O professor responsável pela disciplina tem total autonomia para deferir ou indeferir tais pedidos . As inscrições em disciplina isolada deferidas , para posterior emissão de declaração e controle . Aluno e tutor se reúnem para verificar o histórico do aluno , suas necessidades e sugerir alterações em sua inscrição , visando o melhor aproveitamento do curso . Os alunos para confirmar a matrícula nas disciplinas e encaminha as mudanças à secretária . Alunos de mestrado precisam da assinatura do orientador no requerimento .

O aluno também pode no período pré-determinado pela UNIRIO (PROGRAD) trancar a matrícula no curso .

O trancamento é de no máximo 1 semestre e não pode ser feito no primeiro semestre .

As respostas do questionário estão resumidas na Tabela 7. Um número selecionado de questões foi avaliado, com enfoque nas opiniões dos participantes sobre as diferenças características do estudo e da ferramenta utilizada.

Tabela 7 – Questionário com a contagem de respostas

Pergunta	Concordo	Indiferente	Discordo
A ferramenta permitiu a expressão de sua visão sobre a inscrição de disciplinas?	10	1	1
Contar a história de forma coletiva facilitou a expressão da sua visão sobre a inscrição em disciplinas?	9	2	1
A funcionalidade de anexar documentos a um evento específico da história foi útil para complementar sua visão do evento?	7	5	0
A detecção automática de eventos sem personagem instigou a criação de novos personagens?	6	4	2
A funcionalidade de votação foi útil para resolver incoerências e conflitos?	2	10	0
A operação de manipulação de eventos “Trocar de Lugar” foi útil para um melhor encadeamento dos eventos?	9	2	1
A operação de manipulação de eventos ”Unir Eventos” foi útil para um melhor encadeamento dos eventos?	9	3	0
A operação de manipulação de eventos “Fragmentar Eventos” foi útil para um melhor encadeamento dos eventos?	9	3	0

A Tabela 7 mostra uma visão concisa do estudo de caso. A maioria dos participantes concordou com os benefícios da narrativa colaborativa para elicitação de processos e com a utilidade das funcionalidades adicionais.

As primeiras perguntas foram muito importantes, já que elas são relacionadas com a validade do método para elicitação de processos e a ferramenta ProcessTeller. Embora a maioria dos usuários tenha confirmado a sua utilidade, um pequeno grupo discordou. Os principais motivos foram: a mistura entre eventos pertencentes a diferentes processos de inscrição em disciplinas na mesma história e a impossibilidade, no escopo do estudo de caso, de cada usuário contar uma história individualmente.

Três perguntas tiveram menos de 75% de aprovação (i.e. 9 respostas em Concordo) dos usuários. As duas primeiras se referem à detecção de personagens e ao anexar de documentos aos eventos. Neste caso, a baixa taxa de aprovação poderia ser explicada pelo reduzido número de novos eventos conforme a história se desenvolveu e o processo descrito atingiu sua completude na narrativa. Com menor número de novos eventos, é admissível que alguns usuários não pudessem ter experimentado a funcionalidade de detecção de personagens, o que poderia explicar o número de usuários em "Discordo" e um número relativamente alto de repostas "Indiferente".

A última pergunta se referia à funcionalidade de Votação e obteve o maior número de respostas "Indiferente" de todo o questionário, assim não foi registrada a ocorrência de nenhuma resposta "Discordo". Isto aponta claramente que a Votação não foi utilizada durante o estudo de caso e que os usuários preferiram utilizar comentários para discutir assuntos relacionados aos eventos.

Além das perguntas de múltipla escolha, outras três perguntas abertas e opcionais estavam presentes. Elas eram relacionadas especificamente a três enfoques: (i) O ponto de vista de cada usuário descrito na história, (ii) A história final gerada e a validade como descrição de um processo e (iii) comentários gerais sobre o processo de elicitação como um todo. Algumas tendências específicas dos usuários foram observadas, com base nas suas respostas individuais a estas perguntas.

Muitos usuários se disseram incapazes de afirmar que a história refletia o processo de Inscrição em Disciplinas de forma completa e correta, devido ao fato de possuírem um ponto de vista limitado sobre o processo como um todo. Mas eles estavam confiantes de que suas contribuições individuais estavam corretas e que suas experiências com o processo em questão estavam refletidas na história.

Outro grupo de usuários expressou surpresa com os detalhes presentes na história, desconhecidos para eles. Na opinião deles, isto se devia ao contato limitado com apenas parte do processo em suas atividades de dia-a-dia.

Outra observação feita por vários dos participantes se refere à abundância de conhecimento, persistindo a crítica de que existe uma mistura de eventos relativa à inscrição em disciplinas do mestrado e da graduação na história final.

Por fim, a maioria concordou que seu ponto de vista, limitado ou não, sobre o processo foi descrito corretamente na história e se declararam satisfeito com a narrativa final, considerada por eles como a união destas pequenas contribuições de uma forma colaborativa.

Outros comentários foram extraídos do questionário, como a proposta de diversas modificações e novas funcionalidades na ferramenta ProcessTeller, desde o uso de estruturas lógicas complexas (como AND, OR, XOR) até melhorias gerais na interface do software.

Em geral, as respostas das perguntas discursivas, e além do fato inicial descrito no estudo de caso, a forma como a contribuição inicial do primeiro usuário acarretou um aumento considerável de contribuições de outros usuários (na forma de novos eventos, comentários, além de personagens e documentos) reforça o valor e a utilidade do elemento colaborativo da técnica de Group Storytelling. Possivelmente as pessoas não iriam relembrar várias partes da história ou não se sentiriam estimuladas para tal, sem o auxílio dos elementos narrativos do primeiro usuário.

Outro fato inesperado foi o aumento expressivo dos comentários registrados nos eventos, principalmente conforme a história aumentou de tamanho e se tornou mais "madura" no sentido de expressar mais fielmente todos os pontos de vista sobre o processo dos narradores. Isto evidencia claramente um foco maior em formas de colaboração adicionais, como fóruns e chats, de modo a auxiliar o processo de elicitação como um todo.

7.3. Estudo de Caso 2: Processo "Gerir Levantamento de Processos"

7.3.1. Descrição

O segundo estudo de caso foi realizado com a equipe de Processos de uma organização brasileira de grande porte, cujo nome está omitido neste texto por

razões de confidencialidade Foi selecionado um processo intensivo em conhecimento denominado “Gerir Levantamento de Processos”.

Os narradores candidatos foram selecionados dentre membros da equipe que participam da atividade de modelagem de processos, membros que tem pouco contato com esta atividade (realizam tarefas operacionais), e a gerente da equipe. Como no estudo de caso anterior, a seleção dos narradores é uma tentativa de uma maior abrangência em termos de variados pontos de vista sobre o mesmo processo de negócio.

Após esta seleção, foi enviado um convite para todos os membros da equipe que desejassem participar e uma apresentação oral, com uma demonstração dos conceitos principais de Group Storytelling, o método Story Mining e a ferramenta ProcessTeller foi feita para todos os participantes.

Esta apresentação, realizada nas instalações da própria organização, terminou com a criação de uma história exemplo, onde os participantes deveriam contar o que havia ocorrido durante a apresentação, com orientação presencial dos apresentadores, de modo que a familiaridade com a ferramenta e os conceitos do processo colaborativo de contar histórias não se tornassem obstáculos para o processo de elicitação do processo de negócio alvo.

O grupo de narradores então ficou definido em 8 pessoas, compondo o total da equipe envolvida no processo de negócio que concordou em participar no estudo.

O estudo de caso foi realizado por três semanas, em paralelo com as atividades gerais da equipe envolvida. Vários eventos foram cadastrados, sendo que alguns deles continham grandes quantidades de textos e de detalhes sobre as atividades descritas. Apenas um comentário foi registrado e este foi uma advertência da gerente da equipe para que um dos narradores descrevesse alguns elementos específicos em um evento previamente registrado.

Cinco personagens foram cadastrados no total e alguns deles tinham nomes com siglas, que possivelmente apenas os envolvidos no processo ou na organização poderiam conhecer. As informações complementares dos personagens conseguiram solucionar com sucesso esta falta de informação e um dos campos, “Relacionamento interpessoal”, foi utilizado para descrever a interação do personagem com o processo como um todo.

Nenhum documento foi cadastrado pelos participantes, o que tornou a história e seus elementos a única fonte de informação sobre o processo a ser modelado.

Com fim do período estipulado para a criação da história, a segunda fase do método teve início. Por causa da ausência de documentos, o conteúdo das histórias foi utilizado sem a filtragem por termos. Após a execução do workflow contendo as técnicas de Processamento da Linguagem Natural, durante a segunda fase do método, 20 atividades foram extraídas.

7.3.2. Análise

A análise preliminar das atividades extraídas constatou que mesmo sem os documentos para filtrar possíveis trechos irrelevantes das histórias, todas as atividades extraídas foram relacionadas ao processo a ser modelado. A existência de apenas um comentário, cujo conteúdo é relacionado a uma mensagem entre gerente e membro de equipe, para a modificação de um dos eventos cadastrados, demonstra dinâmica mais "fechada", centrada em um único núcleo de trabalho.

Uma comparação foi realizada entre os um modelo de processo de negócio já existente na organização, de onde foram retiradas para análise as atividades referentes ao processo "Gerenciar Levantamento de Processos" (PM1) com as atividades do proto-modelo gerado pelo método Story Mining (PM2). A Tabela 8 mostra um quadro comparativo entre as atividades extraídas nos dois modelos.

Tabela 8 – Resultado comparativo entre os dois modelos (Gramática I)

Estatística	PM1	PM2
Total de atividades	18	20
# de atividades coincidentes	6	8
# de atividades não-coincidentes	12	12

Um dos primeiros pontos interessantes que surgiram na análise entre os modelos foi o uso de termos diferentes para uma mesma definição. Embora no modelo pré-existente exista a menção a duas reuniões iniciais, uma de alinhamento e outra de prospecção, o processo descrito pela narrativa descreve uma reunião de "kick-off".

Após este evento temos uma necessidade de reunião de alinhamento descrita por outro narrador, levando à conclusão de que a reunião de "kick-off" possui o

significado de reunião de prospecção, o que foi confirmado após uma análise mais minuciosa.

Outro detalhe interessante foi que a reunião de alinhamento foi apenas mencionada e não descrita em profundidade, o que poderia indicar uma menor importância ou algo realizado de forma "ad-hoc" pelos membros da equipe.

Durante toda a narrativa, podemos perceber que as reuniões são um dos focos principais da história, possivelmente por se tratar de um dos momentos mais marcantes para toda equipe, onde são tomadas decisões estratégicas e se produz a documentação sobre as atividades posteriores.

Outros elementos, como detalhes técnicos sobre a ferramenta utilizada (ARIS Business Modeler) nas atividades também estavam presentes no texto, indicando partes do processo que são específicos da equipe e não se encontravam presentes no modelo pré-existente.

A diferença entre o nível de detalhe de algumas atividades é significativa, como é o caso da descrição da preparação da agenda de entrevistas, inexistente no modelo pré-existente e no foco no alinhamento à iniciativa estratégica de negócio da Área, que foi descrito pela equipe, mas não foi descrito no modelo previamente criado do mesmo processo.

O enfoque diferente é ainda mais evidente se compararmos de uma forma geral as atividades dos dois modelos, encontrando divergências como a inexistência de atividades anteriores às reuniões no modelo gerado, que tratam da "intenção de modelagem de processos" sendo encaminhada e analisada pela equipe que contou as histórias. Isto poderia indicar uma parte do processo que raramente é realizada ou algo tido como de menor importância pelos participantes. De acordo com a narrativa dos participantes, o cliente expõe suas necessidades durante a reunião de "kick-off", o que poderia ser um ponto de vista real, fruto da experiência prática da equipe durante a execução de uma instância do processo.

A ausência de algumas atividades do modelo previamente construído, referente à gestão do levantamento por parte do administrador de processos de negócio, também reforça a evidência de que a narrativa teve como foco principal as

atividades nas quais os participantes possuem um envolvimento direto e não na gestão de levantamento de processo como um todo.

Finalmente, uma descrição concisa de auditoria sendo realizadas a cada 15 dias nos modelos de processos criados pela equipe, possibilitou a extração de elementos que não se encontram na parte de Gerência de Levantamento de Processos, no modelo anterior. Pode-se atribuir este fato à presença destas atividades em outro macro-processo deste modelo ou como informação nova sobre o processo alvo.

As tabelas a seguir descrevem a análise comparativa entre as atividades dos dois modelos de processos.

Tabela 9 – Atividades do Processo validado (Gerência de Levantamento de Processos)

Processo Validado	
Item	Descrição do Item
5.2.1.1	Informar Intenção de modelagem de processos
5.2.1.2	Encaminhar intenção de modelagem de processos
5.2.1.3	Analisar intenção de de modelagem de processos
5.2.1.4	Notificar não aprovação de modelagem de processos
5.2.1.5	Notificar não aprovação de modelagem ao solicitante
5.2.1.6	Informar demanda de modelagem de processos
5.2.1.7	Monitorar execução do projeto de modelagem
5.2.1.8	Realizar reunião de alinhamento da execução do projeto
5.2.1.9	Realizar reunião de prospecção do projeto
5.2.1.10	Enviar template de documentação de planejamento do projeto
5.2.1.11	Preencher documentação de planejamento do projeto
5.2.1.12	Analisar documentação de planejamento do projeto
5.2.1.13	Solicitar correção de documentação de planejamento do projeto
5.2.1.14	Corrigir documentação de planejamento do projeto
5.2.1.15	Avaliar proposta de modelagem de processos
5.2.1.16	Notificar não-aprovação de proposta modelagem de processos
5.2.1.17	Consolidar objetivos candidatos e globais
5.2.1.18	Participar da reunião de validação do processo

Tabela 10 – Atividades extraídas (Gramática I) relacionadas com o modelo validado

Atividade Extraída	Atividade do modelo validado
O gerente de projeto envia email para o cliente para agendamento da reunião de kick-off	5.2.1.9
Ë necessária uma reunião para alinhar os objetivos	5.2.1.8
O gerente de projeto prepara juntamente com a equipe	5.2.1.9
O Escritório de Processos envia ao cliente toda a documentação existente sobre as regras para modelagem no repositório de processos de Do Escritório de Processos apresenta como é a gestão de um projeto de modelagem com as etapas de criação de infraestrutura	5.2.1.10 5.2.1.9
reunião resulta a lista dos especialistas a serem entrevistados para levantamento das informações	5.2.1.9
caso o Administrador julgue	5.2.1.15, 5.2.1.12
que o mesmo não está alinhado à iniciativa estratégica de negócio	5.2.1.16

Tabela 11– Atividades novas extraídas pelo método (Gramática I)

Atividade Extraída	Atividade do modelo validado
O cliente expõe a sua necessidade	Nova
O gerente de projeto prepara agenda de entrevistas	Nova
do projeto deverá levantar as informações para estabelecer o escopo	Nova

Isto implica na definição dos macro-processos	Nova
e processos a serem modelados	Nova
O Escritório de Processos solicita para Administrador a liberação	Nova
do projeto não pode ser iniciada	Nova
Área. Administrador autoriza a modelagem de processos no repositório corporativo de	Nova
E a equipe de modelagem iniciar o projeto	Nova
do os processos seguindo os padrões de modelagem de processos	Nova
O Escritório de Processos realiza a cada 15 dias uma auditoria básica no projeto que está sendo modelado	Nova
São executados os scripts do ARIS com categoria Auditoria	Nova

Tabela 12 – Atividades extraídas com o uso da Gramática II

Atividade Extraída
O gerente de projeto envia email para o cliente para agendamento da reunião de kick-off.
É necessária uma reunião para alinhar os objetivos do projeto.
O gerente de projeto prepara juntamente com a equipe do Escritório de Processos os documentos para a reunião.
O Escritório de Processos envia ao cliente toda a documentação existente sobre as regras para modelagem no repositório de processos da Área. São elas : Instruções para Modelagem de Processos , Diretrizes de Modelagem de Processos da Área . O

cliente expõe a sua necessidade .
A equipe do Escritório de Processos apresenta como é a gestão de um projeto de modelagem com as etapas de criação da infraestrutura (
Reunião resulta a lista dos especialistas a serem entrevistados para levantamento das informações .
O gerente de projeto prepara agenda de entrevistas e a envia para o cliente .
Nesta etapa a equipe de projeto deverá levantar as informações para estabelecer o escopo e objetivos do projeto .
Isto implica na definição dos macro - processos e processos a serem modelados .
O Escritório de Processos solicita para Administrador a liberação do projeto
A modelagem do projeto não pode ser iniciada caso o Administrador julgue que o mesmo não está alinhado à iniciativa estratégica de negócio da Área . Administrador autoriza a modelagem de processos no repositório corporativo da Área . O Escritório de Processos disponibliza os ambientes necessários para a equipe de modelagem iniciar o projeto .
Os processos seguindo os padrões de modelagem de processos e de estruturas de grupos de diretórios do repositório ARIS . O Escritório de Processos realiza a cada 15 dias uma auditoria básica no projeto que está sendo modelado . São executados os scripts do ARIS com categoria Auditoria e reportado ao cliente inconformidades com os padrões de modelagem e de estrutura de grupos .

O número reduzido de eventos na história, mesmo com um processo-alvo abrangente como “Gerir Levantamento de Processos”, é um fato que surpreende, já que se esperava um número maior de eventos em vista do grau de “abstração” de tal processo de negócio. No entanto, o teor bastante claro e direto das narrativas, em vista de que os envolvidos são da mesma equipe e trabalham com a área de

modelagem e elicitação, mostra que praticamente todos os eventos continham conhecimento relevante e quase nada que pudesse ser descartado.

O uso de comentários como forma de interação entre participantes se manteve, mesmo que muito reduzido em comparação ao estudo anterior. Considerando que os participantes deste estudo de caso se encontravam fisicamente no mesmo espaço físico durante as três semanas, esta se torna mais uma evidência de que os comentários são parte integrante da construção colaborativa das narrativas.

O registro de personagens com siglas demonstra o aparecimento de uma "linguagem própria" dentre os participantes, em vista do domínio a ser trabalhado e a própria cultura organizacional e os hábitos dentre de uma equipe. Embora as informações complementares dos personagens tenham resolvido este problema, ele ressalta uma importância maior das informações de contexto que possivelmente estariam presentes também em documentos relativos ao processo, os quais não foram cadastrados pela ferramenta.

Tabela 13 – Questionário com a contagem de respostas

Pergunta	Concordo	Indiferente	Discordo
A ferramenta permitiu a expressão de sua visão sobre a inscrição de disciplinas?	4	0	0
Contar a história de forma coletiva facilitou a expressão da sua visão sobre a inscrição em disciplinas?	4	0	0
A funcionalidade de anexar documentos a um evento específico da história foi útil para complementar sua visão do evento?	4	0	0
A detecção automática de eventos sem personagem instigou a criação de novos personagens?	3	0	1
A funcionalidade de votação foi útil para resolver incoerências e conflitos?	3	1	0
A operação de manipulação de eventos "Trocar de Lugar" foi útil para um melhor encadeamento dos eventos?	4	0	0

A operação de manipulação de eventos ”Unir Eventos” foi útil para um melhor encadeamento dos eventos?	4	0	0
A operação de manipulação de eventos “Fragmentar Eventos” foi útil para um melhor encadeamento dos eventos?	4	0	0

O questionário obteve resposta de apenas metade dos participantes, portanto podemos considerar apenas algumas evidências dele. Foi apontado novamente que as votações não foram utilizadas e não tiveram qualquer relevância no processo. Novamente, como no estudo de caso anterior, existe resistência à detecção automática de personagens na inclusão de novos eventos.

Outro ponto interessante, veio das perguntas discursivas. Um participante afirmou que o método pode funcionar muito bem, no entanto, apenas para processos pequenos, pois poderia se tornar confuso, talvez devido à grande quantidade de eventos a serem inseridos. Outro participante fez uma crítica parecida, dizendo que não conseguia vislumbrar um consenso rápido se o estudo envolvesse uma grande quantidade de participantes, o que poderia levar a várias reuniões de grupo. Isto pode indicar uma necessidade de formas de colaboração adicionais ao método, de modo a estimular a colaboração e interação entre os narradores.

7.4. Discussão

Os dois estudos de caso em questão foram realizados em cenários diferentes. O primeiro em um ambiente universitário e de livre-acesso ao conhecimento, com um processo de negócio a ser modelado de conhecimento geral. Já no segundo caso de estudo, foi proposta a elicitação de um processo de negócio de maior abstração e de nível gerencial, ou seja, não é de conhecimento geral de todos os funcionários da organização.

O proto-modelo gerado em ambos os estudos de caso apresentou um grau de “ruído” (i.e. informações indesejadas para o objetivo do estudo), principalmente na forma de algumas atividades incompletas (Ex: ”Podem ser cursadas até duas disciplinas por semestre”) ou atores incorretos (Ex: “Grade de horários”). Este tipo de ocorrência,

típico de aplicações de Mineração de Texto, indica uma maior atenção para técnicas mais refinadas da área, visando resultados mais precisos.

Uma análise comparativa pode ser feita entre os dois casos, apesar de estarem em contextos diferentes, tais como, grande organização x pequena organização e ambiente universitário x ambiente empresarial.

Pode-se perceber que o fluxo de conhecimento, mensurável na quantidade de comentários e manipulações de eventos, parece ser mais intenso no primeiro estudo de caso. Poderia se atribuir a isso a maior liberdade entre os participantes e ausência de hierarquia e subordinação. O segundo estudo, focado em uma empresa de grande porte e em uma equipe específica, demonstrou, pela ausência de comentários e pela natureza bastante direta dos eventos das histórias, uma possível interação externa à ferramenta para discussão do que seria narrado.

Mesmo assim, houve um comentário na história do segundo estudo de caso, evidenciando sua importância para a construção colaborativa das histórias. Um suporte maior ao fator colaborativo entre usuário poderia ser dado em novas versões da ferramenta ProcessTeller, possivelmente com uma funcionalidade de fórum ou chat.

Muitos participantes sugeriram, pelas respostas dos questionários, diversas modificações da ferramenta ProcessTeller. Desde encadeamentos mais complexos entre eventos (utilizando estruturas lógicas como AND, OR e XOR) e um maior suporte à colaboração como o uso de ferramentas colaborativas adicionais como chats e fóruns.

O uso intensivo dos comentários mostra que uma maior atenção deve ser dada, em pesquisas futuras, para formas de mineração de texto e processamento da linguagem natural de elementos auxiliares da história, ampliando o foco da mineração para abranger, por exemplo, os comentários das histórias.

Quanto à mineração, constatou-se uma diferença de resultados entre as duas gramáticas. A primeira, voltada a uma estruturação detalhada do texto, detalhando palavras que possam indicar fluxos, embora seja computacionalmente mais adequada, foi percebido que em suas atividades de texto, muitas descrições estão incompletas, provocando a necessidade de o analista voltar à história original para conferir todo o contexto de uma atividade incompleta.

A segunda gramática, possuindo um foco mais geral, extraiu atividades mais completas, possuindo como única falha ocasional o surgimento de mais de uma atividade do processo no texto de cada atividade minerada.

A comparação entre as gramáticas foi possível principalmente pela utilização de um SGWfC na implementação do método, permitindo uma flexibilização entre as técnicas utilizadas. As duas formas de extrair atividades (orientada a estrutura e orientada à extração geral) podem ser utilizadas em conjunto, possibilitando formas de mineração com objetivos diferentes em um mesmo repositório.

A geração de dois tipos de saída foi bastante útil para avaliar a implementação, já que foi possível comparar os trechos de texto adquirido a partir do template utilizado e as atividades e raias do modelo BPMN gerado.

Os resultados da mineração extraíram diversos elementos relacionados ao processo, como regras de negócio e os inseriu como atividades e atores. Isto se deve à liberdade dada aos narradores de expor seu conhecimento sobre o processo.

No entanto, isto pode ser resolvido na edição final do modelo e o simples fato do método capturar estes importantes fragmentos de conhecimento sobre o processo de negócio, mesmo que representados como outros elementos de processo, reforça as vantagens da mineração em narrativas.

A aplicação bem-sucedida do método e a descoberta de conhecimento novo, que não estava presente nos modelos existentes nas organizações envolvidas nos estudos de caso desta dissertação, demonstram a aplicabilidade do Story Mining em diversos perfis de processos e organizações. Desta forma, o método, embora não se apresente como solução ideal nem como substituição das técnicas "tradicionais" de elicitação e modelagem de processos, se mostra como uma ferramenta valiosa, passível de aplicação conforme a necessidade do analista, do processo e da organização envolvida em cada caso.

Capítulo 8 – Conclusões e Trabalhos Futuros

A área de Modelagem de Processos de Negócio possui grande importância nas organizações atuais, sendo fator crítico para análise e compreensão de como seus membros atuam e como a organização como um todo funciona. Um dos grandes problemas desta tarefa reside na sua fase de Elicitação, onde se propõe a aquisição de conhecimento sobre o processo a ser modelado, geralmente a partir das pessoas que participam de suas atividades ou tem contato com as tarefas envolvidas. Esta não é uma tarefa trivial, devido ao fato de que, em geral, o conhecimento do processo é fruto das experiências pessoais de cada indivíduo e difícil de expressar de forma clara e concisa. Mais ainda, o conhecimento sobre o processo geralmente se encontra fragmentado, cada pessoa, em geral, conhecendo apenas a parte que tem contato no dia-a-dia de trabalho.

A pesquisa realizada nesta dissertação abordou o problema de aquisição de conhecimento para modelagem de processos de negócio, estabelecendo inicialmente os problemas a serem analisados, buscando soluções em diferentes áreas do conhecimento.

A proposta do uso de narrativas colaborativas do Group Storytelling, para permitir que cada participante do processo possa expressar seu conhecimento livremente e de técnicas de Mineração de Texto para fornecer suporte à interpretação deste conhecimento expresso foi selecionada como uma solução adequada para os problemas descritos.

A importante questão do viés do analista, e sua influência na qualidade do modelo final, produto das tarefas de elicitação e modelagem de processos, também foi analisada. Pelos estudos de caso, pode-se perceber que o viés do analista foi reduzido e novo conhecimento a respeito do processo-alvo, que não se encontrava nos modelos previamente existentes nas organizações, foi encontrado pelo método.

Este fato pode ser considerado como uma forte indicação de que a redução do viés pode ser um fator importante para a qualidade do modelo final e a fidelidade da representação

do processo de negócio, embora estudos futuros mais extensos sejam necessários para comprovar esta afirmação.

Seguiu-se à proposta do método Story Mining, sua implementação e a avaliação de resultados, visando estudar sua viabilidade como forma de elicitação de processos de negócio. Esta avaliação se fez por dois estudos de caso com perfis propositadamente diferentes: O primeiro, em um ambiente acadêmico, em uma cultura organizacional mais livre e o segundo em um ambiente empresarial, com regras e normas a serem obedecidas. A análise dos dois "extremos", no que se refere ao resultado da atividade de elicitação de um processo de negócio utilizando o método StoryMining, trouxe evidências sobre como o método funcionaria em casos diferentes e comprovou a viabilidade e utilidade do método para a área de Modelagem de Processos de Negócio, principalmente para a tarefa de Elicitação de Processos de Negócio ou Levantamento de Informações.

Por fim, pode-se constatar que os requisitos iniciais para a proposta do método foram validados. A utilização de narrativas (Requisito A) e o enfoque no aspecto colaborativo da elicitação (Requisito B) foram alcançados com o uso da técnica Group Storytelling. Isto permitiu que cada pessoa expressasse livremente sua visão sobre o processo (Requisito C), mitigando os possíveis efeitos negativos do viés de um analista (Requisito D), em uma técnica tradicional, por exemplo.

A aplicação de técnicas de Mineração de Texto para suporte à análise de narrativas e geração de modelos atingiu os objetivos desejados (Requisito E), permitindo que novo conhecimento fosse extraído e apresentado em uma forma mais próxima da linguagem própria do analista de processos.

8.1. Considerações Finais e Trabalhos Futuros

Algumas considerações podem ser feitas sobre o trabalho de pesquisa em si, principalmente em vista das evidências extraídas, tanto de cada estudo de caso, como da comparação entre os resultados de ambos e as experiências dos participantes, expressas pelas respostas dos questionários e comentários da história.

O uso de mineração de texto em uma história torna possível a extensão da pesquisa para uma aplicação de maior escala, possivelmente abordando um repositório de histórias sobre processo e atividades realizadas pela organização e sua posterior exploração utilizando técnicas de Mineração de Texto, de forma a extrair não apenas o fluxo de

atividades dos processos mas também regras de negócio, informações de contexto e outros tipos de conhecimento presentes nas narrativas colaborativas. De fato, o conhecimento adquirido durante a execução dos estudos de casos da dissertação não se restringiram às atividades de um processo, mas informações extras foram encontradas, como regras de negócio e detalhes previamente desconhecidos sobre o processo como um todo, incluindo informações relevantes sobre seu contexto. Esta é uma direção de pesquisa que pode ser mais explorada.

Trabalhos futuros podem incluir também a aquisição automática de outros elementos de um modelo de processo de negócio, incluindo eventos de um processo, fluxos de ações, dentre outras partes que formam um modelo de processo representativo. Além disso, técnicas de Mineração de Texto mais refinadas, com o possível uso de outras fontes de conhecimento, também podem ser exploradas, visando diminuir o grau de ruído durante a segunda fase do método Story Mining.

O uso de SGWfC como suporte para execução do método também mostrou vantagens por facilitar a definição do workflow, e a comparação de resultados entre instâncias distintas. As primeiras execuções do método, utilizando scripts e posteriormente classes desenvolvidas em linguagem Python não possuíam o nível de flexibilidade, muito útil para o trabalho experimental com diferentes técnicas e gramáticas que o uso de um SGWfC proporcionou.

A utilização de duas gramáticas mostrou uma conclusão interessante. Embora a primeira gramática, mais detalhada, permita uma estruturação mais rica a partir do texto livre, a segunda gramática extraiu atividades completas, mesmo sem registrar palavras que indiquem fluxos de ações e outros elementos mais complexos de processos de negócio. Isto indica que para o suporte do analista de processos, talvez uma abordagem mais simples e menos estruturada seja o caminho a ser seguido.

A livre-expressão de conhecimento, um dos pontos fortes identificados na técnica Group Storytelling, foi alvo de sugestões por partes dos usuários do ProcessTeller nos dois estudos realizados. Muitas pessoas sugeriram o uso de mais de uma história para um mesmo processo e a ampliação das formas de encadeamento entre eventos, inclusive relações lógicas complexas como AND, OR e XOR, por exemplo.

Outra ponto importante a ser considerado para uma possível extensão do método é o formato do proto-modelo gerado. Diversos formatos de modelagem de processo poderiam ser utilizados, de acordo com a aplicação em vista. Outras formas de

representação independentes da modelagem de processos poderiam também ser geradas, utilizando-se o workflow científico como base e a customização das gramáticas de expressões regulares como forma de definir *templates* para atividade EI, poderia se extrair não apenas processos, mas também estruturas conceituais como ontologias.

O método parece ser adequado para uso em modelagem de processos de negócio onde se possua pouca ou nenhuma informação disponível sobre o processo em si. O uso de estudos de casos, com comparação de modelos de processos pré-existentes obtidos através de técnicas conhecidas como entrevistas, indica também o uso possível do Story Mining para refinar e aprimorar processos de negócio já conhecidos.

As técnicas de Mineração de Texto utilizadas podem ser aprimoradas, inclusive com o tratamento de co-referência, bastante comum em linguagem informal, como o texto de uma narrativa ou história. Outras possíveis melhorias são utilizar técnicas mais refinadas para etiquetação sintática e morfológica, de modo a reduzir o erro na classificação de palavras e, consequentemente, melhorar o grau de qualidade da extração de atividades do processo.

8.2. Contribuições Realizadas

A contribuição principal da pesquisa foi a proposta de um método para modelagem e elicitação de processos de negócio, que visa mitigar os obstáculos em nível de conhecimento para atingir um modelo de qualidade. Com exceção de poucos trabalhos que indicam esta direção de pesquisa, referenciados nesta dissertação, não foi encontrado nenhum outro trabalho na literatura que especificasse um método a ser utilizado para a união das áreas de Mineração de Texto com a Modelagem de Processos de Negócio com a geração de um modelo passível de edição, nem seus estudos de casos e implementação, semelhante ao trabalho de pesquisa desta dissertação.

Além disso, podemos destacar outras contribuições resultantes do trabalho:

- A realização de uma revisão bibliográfica da literatura de Modelagem de Processos de Negócio, Elicitação de Conhecimento e Elicitação de Requisitos de Software, permitindo a identificação dos problemas relevantes sobre o conhecimento envolvido e sua aquisição.
- O desenvolvimento de uma nova ferramenta para Group Storytelling orientada a Processos de Negócio, denominada ProcessTeller.

- A adaptação de um modelo de cenários para uso com os possíveis fluxos de atividades de um modelo de processo de negócio.
- O desenvolvimento de módulos específicos para utilização de técnicas de mineração de texto com ferramentas conhecidas em um workflow científico.
- Os resultados dos estudos de casos, apresentando evidências sobre a modelagem de processos em cenários diferentes, a partir da expressão de conhecimento pelos usuários, reduzindo o papel do analista ao de moderador.

Referências Bibliográficas

AALST, W.M.P., WEIJTERS, A., 2005, *Process-Aware Information Systems: Bridging People and Software through Process Technology*, Wiley & Sons.

ABECKER, A., DECKER, S., 1999, "Organizational Memory: Knowledge Acquisition, Integration and Retrieval Issues" In: Proceedings of XPS-99. Springer LNAI 1570, pp.113-124.

ACHOUR, C. B., 1998, "Guiding Scenario Authoring". In: Proc. 8th European-Japanese Conference on Information Modelling and Knowledge Bases, Finland, pp. 152-171.

ACKERMAN, M.S., 1994, "Augmenting the Organizational Memory: A Field Study of Answer Garden" In: Proceedings of CSCW'94. New York: The Association for Computing Machinery, pp. 243-252.

AGUILAR-SAVEN, R. S., 2002, "Business process modeling: Review and framework", *International Journal of Production Economics, Vol. 20*, pp. 129-149.

ALUISIO S., PINHEIRO G. M., MANFRIM A.M.P *et al.*, 2004, "The Lácio-Web: Corpora and Tools to advance Brazilian Portuguese Language Investigations and Computational Linguistic Tools" In: Proceedings of LREC 2004, pp. 1779-1782.

ALVAREZ R., 2002, "Discourse Analysis of Requirements and Knowledge Elicitation Interviews" In: Proceedings of 35th Hawaii International Conference on System Sciences, p. 255.

BAEZA-YATES, R., RIBEIRO-NETO, B., 1999, *Modern Information Retrieval*. Addison-Wesley Longman.

BECKER, J., DELFMANN, P., KNACKSTEDT, R. *et al.*, 2002 "KonfigurativeReferenzmodellierung", *Wissensmanagement mit Referenzmodellen. Konzepte für die Anwendungssystem- und Organisationsgestaltung*, pp. 25-144.

BECKER, J., DELFMANN, P., DREILING A. *et al.*, 2004, "Configurative Process Modeling – Outlining an Approach to increased Business Process Model Usability" In: Proceedings of the 15th Information Resources Management Association International Conference, pp. 615-619.

BIRD S., LOPER E., 2004, "NLTk: The Natural Language Toolkit". In: Proc. the ACL demonstration session.

BIRD, S., KLEIN, E., LOPER, E., 2009, *Natural Language Processing in Python*. Disponível em http://www.nltk.org/book. Último acesso em Abril de 2010.

BIEMANN C., QUASTHOFF U., BÖHM K. *et al.*, 2003, "Automatic Discovery and Aggregation of Compound Names for the Use in Knowledge Representations" *J.UCS (Journal of Universal Computer Science)*, vol. 9, n. 6,, pp. 530-541.

BLAKE, C., PRATT, W., 2001, "Better Rules, Few Features: A Semantic Approach to Selecting Features from Text" In: Proceedings of the 2001 IEEE International Conference on Data Mining, pp. 59-66.

BOSILJ-VUKSIC V., 2006, "Business Process Modelling: A Foundation for Knowledge Management", *Journal of Information and Organizational Sciences, v. 30, n. 2*.

BPMN – Business Process Modeling Notation, http://www.bpmn.org/Documents. Último acesso em Janeiro de 2010.

*BROWN CORPUS - A Standard Corpus of Present-Day Edited American English, for use with Digital Computers***,** 1964. Compiled by W. N. Francis and H. Kučera. Brown University. Providence, Rhode Island.

BRILL E., 1992 "A Simple Rule-Based Part-of-Speech Tagger". In: Proceedings of the 3rd conference on Applied natural language processing, pp. 152-155.

CALLAHAN, S. P., FREIRE, J., SANTOS, E. *et al.*, "VisTrails: visualization meets data management" In: Proceedings of the 2006 ACM SIGMOD, New York, NY, pp. 745-747.

CAVALCANTI, M. C., TARGINO, R., BAIÃO, F. *et al.*, 2005, "Managing structural genomic workflows using Web services", *Data and Knowledge Engineering 53 (2005)*, pp. 45–74

CHAVES, A., RINO, L., 2008, "The Mitkov Algorithm for Anaphora Resolution in Portuguese". In: Proceedings of the 8th International Computational Processing of the Portuguese Language (PROPOR 2008), Vol. 5190, Springer Verlag, pp. 51-60.

CHINCHOR, N., 1997, *MUC-7 named entity task definition*, http://www.nlpir.nist.gov/related_projects/muc/proceedings/ne_task.html.

CUNNINGHAM, H., 2005. "Information Extraction, Automatic". In: Encyclopedia of Language and Linguistics (2nd ed.).

CUTTING, D., KUBIEC, J., PEDERSEN, J. *et al.*, 1992 "A Practical Part-of-Speech Tagger", In: Proceedings of the 3rd conference on Applied natural language processing, pp. 133-140.

DAVENPORT, T., PRUSAK, L., 1998, *Working Knowledge: How Organizations Manage What They Know*, Harvard Business School Press.

DAVEY B., COPE C., 2008, "Requirements Elicitation – What's Missing?" In: Issues in Informing Science and Information Technology.

DAVIS A., DIESTE O., HICKEY A., 2006, "Effectiveness of Requirements Elicitation Techniques: Empirical Results Derived from a Systematic Review" In: Proceedings of the 14th IEEE International Requirements Engineering Conference (RE'06).

DE VREEDE, G.-J., SOL, H. G., DICKSON, G. W., 1995 "A group problem solvindo approach to Business Process Redesign: combating organized crime in Amsterdam" In: Proceedings of the 28th Annual Hawaii International Conference on System Sciences – HICSS, pp. 493-502.

DIESTE O., LOPEZ M., RAMOS F., 2008, "Updating a Systematic Review about Selection of Software Requirements Elicitation Techniques" In: Proceedings of the 11th Workshop on Requirements Engineering.

DOLAMIC, L., SAVOY, J., 2009, “When Stopword Lists Make the Difference”, *Journal of the American Society for Information Science and Technology 61, issue 1*, pp. 200-203.

FELDMAN, R., KLOESGEN, W., ZILBERSTEIN, A., 1997, “Visualization Techniques to Explore Data Mining Results for Document Collections”, In: Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining, pp. 16-23.

FELDMAN, R., FRESKO, M., KINAR, Y. *et al.*, 1998, "Text Mining at the Term Level" In: Proceedings of the 2nd European Symposium on Principles of Data Mining and Knowledge Discovery, pp. 65-73.

FELDMAN, R., SANGER, J., 2007, *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*, Cambridge University Press.

GAIZAUSKAS, R., WILKS, Y., 1998, “Information Extraction: Beyond Document Retrieval**”,** *Journal of Documentation***,** Volume 54, Number 1, pp. 70-105(36).

GARCIA, A. C. B., FERRAZ, I. N., PINTO, F., 2006 “The Role of Domain Ontology in Text Mining Applications: The ADDMiner Project “ In: Proceedings of the 6th IEEE International Conference on Data Mining – Workshops, pp. 34-38.

GOGUEN, J. A., LINDE, C., 1993 “Techniques for Requirements Elicitation” In: Proceedings of the Requirements Engineering Conference ’93, pp. 152-164.

GONÇALVES, J. C. A. R., SANTORO F.M., BAIÃO F.A., 2009, "Business Process Mining from Group Stories”. In: Proc. 13th International Conference on Computer-Supported Cooperative Work in Design., Santiago, Chile.

GONÇALVES, J. C. A. R., SANTORO F. M., BAIÃO F.A., 2010a, " A Case Study on Designing Business Processes Based on Collaborative and Mining Approaches”. In: Proc. 14th International Conference on Computer-Supported Cooperative Work in Design., Shanghai, China.

GONÇALVES, J. C. A. R., SANTORO, F. M., BAIÃO, F. A., 2010b “Collaborative Business Process Elicitation Through Group Storytelling”. In: 12th International Conference on Enterprise Information Systems (ICEIS’2010) – Aceito para publicação.

GONÇALVES, J. C. A. R., SANTORO, F. M., BAIÃO, F. A., 2010c "Let me tell you a story – On how to build process models", *J.UCS (Journal of Universal Computer Science), Special Edition: A selection of the best papers from the 2009 IEEE International Conference on Computer Supported Cooperative Work in Design (CSCWD 2009)* – Aceito para publicação**.**

GHOSE, A., KOLIADIS, G., CHUENG A., 2007, "Process Discovery from Model and Text Artefacts". In: Proc. IEEE Congress on Services.

HICKEY A., DAVIS A., 2004, "A Unified Model of Requirements Elicitation" In: Journal of Management Information Systems, Spring 2004.

HOLZINGER A., GEIERHOFER R., MÖDRITSCHER F. *et al.*, 2008, "Semantic Information in Medical Information Systems: Utilization of Text Mining Techniques to Analyze Medical Diagnoses", *J.UCS (Journal of Universal Computer Science)*, vol. 14, n. 22, pp. 3781-3795.

HOTHO, A., STAAB, S., STUMME, G., 2003, *Text clustering based on background knowledge*, Technical Report 425, University of Karlsruhe, Institute AIFB.

HUANG Y., MITCHELL, T. M., 2006, "Text Clustering with extended user feedback" In: Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 413-420.

INDULSKA, M, RECKER, J., ROSEMANN, M., GREEN, P., 2009, "Business Process Modeling: Current Issues and Future Challenges" In: Proceedings of the 21st International Conference on Advanced Information Systems Engineering, pp. 501-514.

INGVALDSEN J.E., GULLA J.A., SU X. *et al.*, 2005, "A Text Mining Approach to Integrating Business Process Models and Governing Documents". In: Proc. Workshop on Interorganizational Systems and Interoperability of Enterprise Software and Applications.

INGVALDSEN J.E., 2006, "Information Retrieval for Organizational Process Insight". In: Proc. CAISE'06 Doctoral Consortium.

KAISER, K., MIKSCH, S., 2005, *Information Extraction – A Survey*, Technical Report, Asgaard-TR-2005-6, Vienna University of Technology, Institute of Software Technology and Interactive Systems.

KALPIC B., BERNUS, P., 2006, "Business process modeling through the knowledge management perspective", *Journal of Knowledge Management*, vol. 10, n. 3, pp. 40-56.

KODRATOFF, Y., 1999, "Knowledge discovery in texts: A definition and applications.", *Lecture Notes in Computer Science*, 1609, pp. 16–29.

LAPORTI V., BORGES M. R. S., BRAGANHOLO V. P., 2007, "A Collaborative Approach to Requirements Elicitation" In: Proceedings of the 11th International Conference on Computer Supported Cooperative Work in Design.

LAPORTI, V., BORGES, M. R. S., BRAGANHOLO, V. P., 2009 "Athena: A collaborative approach to requirements elicitation" *Computers in Industry 60(6)*, pp. 367-380.

LUTZ, M., 2006, *Programming Python*, O' Reilly Media.

MARTIN, J., 2005, "Fast and Furious Text Mining", *IEEE Data Eng Bull 2005*, 28(4), pp. 11-20.

MITAMURA, T., NYBERG E., 2001 "Automatic Rewriting for Controlled Language Translation", In: Proceedings of the NLPRS 2001 Workshop.

NONAKA, I., TAKEUCHI, H., 1997, *Criação do Conhecimento na Empresa: como as empresas japonesas geram a dinâmica da inovação*, 18a. ed., Editora Campus.

OLIVEIRA, D., 2008, *MiningFlow: Adicionando Semântica a Workflows de Mineração de Texto*, Dissertação de Mestrado, COPPE/UFRJ.

ORENGO V. M., HUYCK C. R., 2001, "A Stemming Algorithm for the Portuguese Language". In: 8th International Symposium on String Processing and Information Retrieval (SPIRE). Laguna de San Raphael, Chile, pp 186-193.

OSBORNE M. "Shallow parsing as part-of-speech tagging,", 2000, In: Proceedings of the 2nd workshop on Learning language in logic and the 4th conference on Computational natural language learning - Volume 7, pp 145-147.

PACHECO, C., TOVAR, E., 2007, "Stakeholder Identification as an Issue in the Improvement of Software Requirements Quality", *Lecture Notes on Computer Science*, 4495/2007, pp. 370-380.

PERRET, R., 2004, "A técnica de Group Storytelling aplicada à gestão de conhecimento", Dissertação de mestrado, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ.

RINDERLE, S., BOBRIK, R., REICHERT, M. *et al*, 2006 "Business Process Visualization – Use Cases, Challenges, Solutions" In: 8th International Conference on Enterprise Information Systems (ICEIS'06), pp 205-219.

ROBERTSON, S., 2004 "Understanding Inverse Document Frequency: On theoretical arguments for IDF", *Journal of Documentation* 60, no. 5, pp. 503-520.

ROLLAND, C., SOUVEYET, C., ACHOUR, C. B., 1998 "Guiding Goal Modeling using Scenarios" In: IEEE Transactions on Software Engineering vol. 24, pp 1055-1071.

SANTORO, F. M., BORGES, M.R.B., PINO, J.A., 2008, "Tell us your process: A Group Storytelling Approach to Cooperative Process Modeling", In: Proc. 12th International Conference on Computer-Supported Cooperative Work in Design. Beijing: IEEE Press, v. I., pp. 29-34.

SANG, E. F. T. K., BUCHHOLZ S., 2000, "Introduction to the CoNLL-2000 Shared Task: Chunking". In: Proc. CoNLL-2000 and LLL-2000, Lisbon, Portugal.

SANG, E. F. T. K., DE MEULDER F., 2003, "Introduction to the CoNLL-2003 shared task: language-independent named entity recognition". In: Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003 - Volume 4, pp. 142-147.

SANTOS, C. K., EVSUKOFF, A. G., LIMA, B. S. L. P. *et al.*, 2009, "Potential collaboration discovery using document clustering and community structure detection" In: Proceedings of 1st ACM international workshop on Complex networks meet information & knowledge management, pp. 39-46.

SINHA A., PARADKAR A., KUMANAN P. *et al.*, 2008, *An Analysis Engine for Dependable Elicitation of Natural Language Use Case Description and its Application to Industrial Use Cases*, IBM Research Report RC24712.

SMITH, E. A., 2001, “The role of tacit and explicit knowledge in the workplace”, Journal of Knowledge Management Vol. 5 (4), pp. 311-321.

TAN, A.-H., 1999, “Text Mining: The State of Art and the Challenges” In: Proceedings of the Pacific Asia Conference on Knowledge Discovery and Data Mining (PAKDD'99) Workshop on Knowledge Discovery from Advanced Databases, pp. 65-70.

TSUJII, J., ANANIADOU, S., 2005 “Thesaurus or Logical Ontology, Which One Do We Need for Text Mining?”, *Computers and the Humanities, Volume 39, Number 1, February 2005* , pp. 77-90.

TSUMAKI T., TAMAI T., 2005, “A Framework for Matching Requirements Engineering Techniques to Project Characteristics and Situation Changes” In: Proceedings of Situational Requirements Engineering Processes (SREP) 2005.

VALLE, C.; PRINZ, W.; BORGES, M.R.S., 2002, “Generation of group storytelling in post-decision implementation process” In: Proceedings of the 7th International Conference on CSCWD 2002, Rio de Janeiro, pp. 361-367.

VALLE, C., RAYBOURN, E. M., PRINZ, W., *et al.*, 2003, “Group storytelling to support tacit knowledge externalization” In: Proceedings of the 10th International Conference on Human - Computer Interaction. Crete, Greece, v. 4, pp. 1218-1222.

WESTON, R.H., CHATHA, K.A., AJAEFOBI, J.O., 2004, “Process thinking in support of system specification and selection”. *Advanced Engineering Informatics*, 18(4), pp. 217-229.

WOODLEY, T., GAGNON, S., 2005, “BPM and SOA: Synergies and Challenges” In: Proceedings of 6th International Conference on Web Information Systems Engineering, LNCS, vol. 3806, pp. 679–688.

XPDL – XML Process Definition Language. http://www.wfmc.org/xpdl.html. Último acesso em Janeiro de 2010.

YANG, Y., PEDERSEN, J.O., 1997, "A Comparative Study on Feature Selection in Text Categorization" In: Proceedings of the 14th International Conference on Machine Learning, pp. 412-420.

YIN, R. K., 2005, *Estudo de caso: planejamento e métodos*, Tradução: Daniel Grassi. 3ª. ed., Editora Brookman, Porto Alegre.

ZHANG Z., 2007, "Effective Requirements Development – A Comparison of Requirements Elicitation Techniques." In: Software Quality Management XV: Software Quality in the Knowledge Society, British Computer Society, pp. 225-240.

Anexo A – Etiquetas do Corpus MAC-MORPHO

Neste anexo, se encontram as etiquetas do corpus MAC-MORPHO para a língua portuguesa, utilizado nesta dissertação. As suas etiquetas complementares não estão listadas aqui, já que o NLTk permitiu que a versão simplificada das etiquetas, apenas tratando de classe gramatical, fosse utilizada.

CLASSE GRAMATICAL	**ETIQUETA**
Adjetivo	ADJ
Advérbio	ADV
Advérbio conectivo subordinativo	ADV-KS
Advérbio relativo subordinativo	ADV-KS-REL
Artigo (definido ou indefinido)	ART
Conjunção coordenativa	KC
Conjunção subordinativa	KS
Interjeição	IN
Nome	N
Nome próprio	NPROP
Numeral	NUM
Particípio	PCP
Palavra denotativa	PDEN
Preposição	PREP
Pronome adjetivo	PROADJ
Pronome conectivo subordinativo	PRO-KS
Pronome pessoal	PROPESS
Pronome relativo conectivo subordinativo	PRO-KS-REL
Pronome substantivo	PROSUB
Verbo	V
Verbo auxiliar	VAUX
Símbolo de moeda corrente	CUR

Anexo B – Gramáticas utilizadas para a Implementação

As duas gramáticas utilizadas na implementação são descritas neste anexo, com base na notação da linguagem de programação Python e do framework NLTk para Processamento de Linguagem Natural.

- Gramática I:

```
NP:   {<ART.*|ADJ.*|NUM.*>+<N.*|ADJ.*|PRO.*|PREP.*|PDEN.*>+}
      {<ART.*><ADV.*>?<V.*|PCP>+<ADV.*>?<N.*|PRO.*|PREP.*|
      PDEN.*|ADJ.*>+}
      {<N.*|PRO.*|PREP.*|PDEN.*|ADJ.*>+}
MISC: {<(|)|:|->+}
VP:   {<ADV.*>?<V.*|PCP>+<ADV.*>?}
PP:   {<PREP.*|PDEN.*>+<NP.*|VP.*>+}
S:    {<NP.*>?<VP><,.*|NP.*|PP.*|VP.*|ADV.*|MISC>+}
      {<NP.*>+<VP>+}
CNJS: {<K.*><,.*>?<S.*|NP.*|VP.*|PP.*>+}
```

- Gramática II:

```
NP:   {<ART.*|ADJ.*|NUM.*>+<N.*|ADJ.*|PRO.*|PREP.*|PDEN.*|K.*>+}
      {<ART.*><ADV.*>?<V.*|PCP>+<ADV.*>?<N.*|PRO.*|PREP.*|
      PDEN.*|ADJ.*|K.*>+}
      {<N.*|!.*|PRO.*|PREP.*|PDEN.*|ADJ.*|K.*>+}
MISC: {<(|)|:|->+}
VP:   {<ADV.*>?<V.*|PCP>+<ADV.*>?}
PP:   {<PREP.*|PDEN.*>+<NP.*|VP.*>+}
S:    {<NP.*>?<VP><,.*|NP.*|PP.*|VP.*|ADV.*|MISC|K.*>+<.>}
      {<NP.*>+<VP>+<.>}
```

Apêndice A – Questionário para os Estudos de Caso

Neste apêndice, temos o questionário utilizado para os Estudos de Caso da Dissertação.

As perguntas e campos com um asterisco (*) são de preenchimento obrigatório.

I. Informações Pessoais

1. Nome Completo: *

2. E-mail: *

3. Login no TellStory: *

II. Questionário

1. A ferramenta permitiu a expressão de sua visão sobre gerir levantamento de processos *

( ) Concordo Totalmente
( ) Concordo
( ) Não fez diferença
( ) Discordo
( ) Discordo Totalmente

2. Contar a história de forma coletiva facilitou a expressão da sua visão sobre gerir o levantamento de processos *

( ) Concordo Totalmente
( ) Concordo
( ) Não fez diferença
( ) Discordo
( ) Discordo Totalmente

3. A funcionalidade de anexar documentos a um evento específico da história foi útil para complementar sua visão do evento *

( ) Concordo Totalmente
( ) Concordo
( ) Não fez diferença
( ) Discordo
( ) Discordo Totalmente

4. A detecção automática de eventos sem personagem instigou a criação de novos personagens *

( ) Concordo Totalmente
( ) Concordo
( ) Não fez diferença
( ) Discordo
( ) Discordo Totalmente

5. A funcionalidade de votação foi útil para resolver incoerências e conflitos *

( ) Concordo Totalmente
( ) Concordo
( ) Não fez diferença
( ) Discordo
( ) Discordo Totalmente

6. A operação de manipulação de eventos “trocar de lugar” foi útil para um melhor encadeamento dos eventos *

( ) Concordo Totalmente
( ) Concordo
( ) Não fez diferença
( ) Discordo
( ) Discordo Totalmente

7. A operação de manipulação de eventos “unir eventos” foi útil para um melhor encadeamento dos eventos *

( ) Concordo Totalmente
( ) Concordo
( ) Não fez diferença
( ) Discordo
( ) Discordo Totalmente

8. A operação de manipulação de eventos “fragmentar eventos” foi útil para um melhor encadeamento dos eventos *

( ) Concordo Totalmente
( ) Concordo
( ) Não fez diferença
( ) Discordo
( ) Discordo Totalmente

9. Releia a história completa escrita pelos participantes e responda: do seu ponto de vista, o processo de “Gerir Levantamento de Processos” está descrito na história criada durante o estudo de caso. Em uma análise mais pessoal, como você percebe sua contribuição para a história? (Pergunta Discursiva) *

10. Insira seus comentários sobre a Ferramenta ProcessTeller: *

11. Insira seus comentários sobre a história gerada: *

12. Insira outras observações gerais (Opcional):

Apêndice B – Resultados dos Estudos de Caso

Neste apêndice, temos o conteúdo dos dois resultados principais do método (O arquivo de "log" das atividades extraídas e o proto-modelo em notação BPMN). Os logs são referentes ao dois estudos de caso e a cada uma das gramáticas aplicadas. O modelo BPMN é referente à segunda gramática, pois esta se mostrou mais coerente para a visualização das atividades.

- Arquivo de Log de Atividades (Estudo de Caso I / Gramática I):

<ATIVIDADE>
<ATOR>O período de inscrição em disciplinas </ATOR>
<ACAO>é definido anualmente </ACAO>
<PARAMETRO>pela direção da UNIRIO </PARAMETRO>
</ATIVIDADE>

<TRIGGER>e</TRIGGER>
<ATIVIDADE>
<ACAO>divulgado </ACAO>
<PARAMETRO>no site oficial </PARAMETRO>
</PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>a grade de horários </ATOR>
<ACAO>enviada </ACAO>
<PARAMETRO>pelo Diretor </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>o aluno </ATOR>
<ACAO>escolhe </ACAO>
<PARAMETRO>as matérias que irá cursar no próximo período </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>O aluno </ATOR>
<ACAO>entra </ACAO>
<PARAMETRO></PARAMETRO>
</ATIVIDADE>

<TRIGGER>e</TRIGGER>
<ATIVIDADE>
<ACAO>faz </ACAO>
<PARAMETRO>uma solicitação de matrícula nas disciplinas </PARAMETRO>
escolhidas </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>A secretária </ATOR>
<ACAO>imprime </ACAO>
<PARAMETRO></PARAMETRO>
</ATIVIDADE>

<TRIGGER>e</TRIGGER>
<ATIVIDADE>
<ACAO>entrega </ACAO>
<PARAMETRO>as solicitações de matrícula dos alunos </PARAMETRO>
aos professores tutores das respectivas turmas </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>Os professores </ATOR>
<ACAO>informam aos </ACAO>
<PARAMETRO>alunos das turmas que são tutores os dias para a confirmação de matrícula </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>as mudanças </ATOR>
<ACAO>enviadas </ACAO>
<PARAMETRO>pelos professores </PARAMETRO>
</ATIVIDADE>

<TRIGGER>e</TRIGGER>
<ATIVIDADE>
<ACAO>confirma </ACAO>
<PARAMETRO>a inscrição dos alunos </PARAMETRO>
</PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>Os professores </ATOR>
<ACAO>entram </ACAO>
<PARAMETRO>em um acordo sobre horários em que gostariam de ministrar suas disicplinas </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>a diretora </ATOR>
<ACAO>organiza </ACAO>
<PARAMETRO>uma grade de horários </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>a diretora </ATOR>
<ACAO>encaminha </ACAO>
<PARAMETRO>a mesmo para a secretaria </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>O horário das disicplinas </ATOR>
<ACAO>é divulgado </ACAO>
<PARAMETRO>no site </PARAMETRO>
</ATIVIDADE>

<TRIGGER>do</TRIGGER>
<ATIVIDADE>
<ATOR>PPGI </ATOR>
<PARAMETRO>O aluno de mestrado </PARAMETRO>
tem </PARAMETRO>
</ATIVIDADE>

<TRIGGER>que</TRIGGER>
<ATIVIDADE>
<ACAO>entregar </ACAO>
<PARAMETRO>o documento de inscrição em disciplinas </PARAMETRO>
</PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>O orientador </ATOR>
<ACAO>assina </ACAO>
<PARAMETRO>o termo de compromisso de orientação para validar sua orientação junto com o aluno de mestrado </PARAMETRO>
</ATIVIDADE>

<TRIGGER>Se</TRIGGER>
<ATIVIDADE>
<ATOR>o aluno </ATOR>
<ACAO>não confirmar </ACAO>
<PARAMETRO>a inscrição em disciplinas com o professor tutor , será considerado abandono de curso por parte </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>O aluno especial </ATOR>
<ACAO>pode escolher cursar </ACAO>
<PARAMETRO>disciplinas isoladas antes de ingressar no curso </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ACAO>Podem ser </ACAO>
<PARAMETRO>cursadas até 2 disciplinas por semestre </PARAMETRO>
, por </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ACAO>Não são </ACAO>
<PARAMETRO>todas </PARAMETRO>
as disciplinas que podem ser cursadas por alunos especiais </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>O aluno especial </ATOR>
<ACAO>deve anexar à </ACAO>
<PARAMETRO>documentação comum para inscrição em disciplinas isoladas uma carta explicando o motivo pelo qual deseja cursar a disciplina </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>O aluno especial </ATOR>
<ACAO>deve comparecer </ACAO>
<PARAMETRO>na secretaria para confirmar </PARAMETRO>
</ATIVIDADE>

<TRIGGER>se</TRIGGER>
<ATIVIDADE>
<ATOR>seu pedido </ATOR>
<ACAO>foi </ACAO>
<PARAMETRO>DEFERIDO </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ACAO>definem </ACAO>
<PARAMETRO>as disciplinas que </PARAMETRO>
serão cursadas no semestre </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>O aluno de mestrado </ATOR>
<ACAO>tem </ACAO>
<PARAMETRO></PARAMETRO>
</ATIVIDADE>

<TRIGGER>que</TRIGGER>
<ATIVIDADE>
<ACAO>escolher </ACAO>
<PARAMETRO>as disciplinas </PARAMETRO>
juntamente com o seu orientador </PARAMETRO>
</ATIVIDADE>

<TRIGGER>do</TRIGGER>

<ATIVIDADE>
<ATOR>aluno </ATOR>
<ACAO>é realizada </ACAO>
<PARAMETRO>no sistema após a entrega </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ACAO>assinado </ACAO>
<PARAMETRO>pelo orientador </PARAMETRO>
</PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>o aluno </ATOR>
<ACAO>ainda tem </ACAO>
<PARAMETRO>um período para inclusão </PARAMETRO>
</ATIVIDADE>

<TRIGGER>do</TRIGGER>
<ATIVIDADE>
<ATOR>PPGI </ATOR>
<ACAO>pode ser feita </ACAO>
<PARAMETRO>num prazo aproximado de 30 dias após o fim </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>O aluno que </ATOR>
<ACAO>trancar </ACAO>
<PARAMETRO>uma disciplina até esse período não é considerado reprovado
</PARAMETRO>
</ATIVIDADE>

<TRIGGER>Se</TRIGGER>

<ATIVIDADE>
<ACAO>não trancar </ACAO>
<PARAMETRO>deve cursar até o fim , </PARAMETRO>
</ATIVIDADE>

<TRIGGER>ou</TRIGGER>
<ATIVIDADE>
<ACAO>será considerado </ACAO>
<PARAMETRO>abandono </PARAMETRO>
</PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>Programa . </ATOR>
<ACAO>Há </ACAO>
<PARAMETRO>um período de discussão de disciplinas a serem ofertadas
</PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>uma grade preliminar </ATOR>
<ACAO>enviada </ACAO>
<PARAMETRO>pela Escola </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>O professor responsável pela disciplina </ATOR>
<ACAO>deve avaliar </ACAO>
<PARAMETRO>o pedido de matrícula em disciplina isolada </PARAMETRO>
</ATIVIDADE>

<TRIGGER>e</TRIGGER>
<ATIVIDADE>

<ACAO>encaminhar </ACAO>
<PARAMETRO>notificação de deferimento </PARAMETRO>
</PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>O professor responsável pela disciplina </ATOR>
<ACAO>tem </ACAO>
<PARAMETRO>total autonomia para deferir </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>as inscrições em disciplina </ATOR>
<ACAO>isolada </ACAO>
<PARAMETRO>deferidas , para posterior emissão de declaração </PARAMETRO>
</ATIVIDADE>

<TRIGGER>e</TRIGGER>
<ATIVIDADE>
<ATOR>tutor se </ATOR>
<ACAO>reúnem </ACAO>
<PARAMETRO>para verificar o histórico </PARAMETRO>
</ATIVIDADE>

<TRIGGER>e</TRIGGER>
<ATIVIDADE>
<ACAO>sugerir </ACAO>
<PARAMETRO>alterações em sua inscrição </PARAMETRO>
, visando o melhor aproveitamento </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>os alunos para </ATOR>

<ACAO>confirmar </ACAO>
<PARAMETRO>a matrícula nas disciplinas </PARAMETRO>
</ATIVIDADE>

<TRIGGER>e</TRIGGER>
<ATIVIDADE>
<ACAO>encaminha </ACAO>
<PARAMETRO>as mudanças à secretária </PARAMETRO>
</PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>Alunos de mestrado </ATOR>
<ACAO>precisam </ACAO>
<PARAMETRO>da assinatura </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>O aluno também </ATOR>
<ACAO>pode </ACAO>
<PARAMETRO>, durante o período pré </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ACAO>determinado </ACAO>
<PARAMETRO>pela UNIRIO </PARAMETRO>
</PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>PROGRAD), </ATOR>
<ACAO>trancar </ACAO>
<PARAMETRO>a matrícula no curso </PARAMETRO>

</ATIVIDADE>

<TRIGGER>do</TRIGGER>
<ATIVIDADE>
<ATOR>o trancamento </ATOR>
<ACAO>é </ACAO>
<PARAMETRO>de </PARAMETRO>
</ATIVIDADE>

<TRIGGER>e</TRIGGER>
<ATIVIDADE>
<ACAO>não pode ser feito </ACAO>
<PARAMETRO>no </PARAMETRO>
primeiro semestre </PARAMETRO>
</ATIVIDADE>

- Arquivo de Log de Atividades (Estudo de Caso I / Gramática II):

<ATIVIDADE>
<ATOR>O período de inscrição em disciplinas </ATOR>
<ACAO>é definido anualmente </ACAO>
<PARAMETRO>pela direção da UNIRIO e divulgado no site oficial . </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>a grade de horários </ATOR>
<ACAO>enviada </ACAO>
<PARAMETRO>pelo Diretor do DIA , o aluno escolhe as matérias que irá cursar no próximo período . </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>

<ATOR>O aluno </ATOR>
<ACAO>entra </ACAO>
<PARAMETRO>no Portal do Aluno e faz uma solicitação de matrícula nas disciplinas escolhidas . </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>O período de inscrição em disciplinas </ATOR>
<ACAO>é definido anualmente </ACAO>
<PARAMETRO>pela direção da UNIRIO e divulgado no site oficial . </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>A secretária </ATOR>
<ACAO>imprime </ACAO>
<PARAMETRO>e entrega as solicitações de matrícula dos alunos aos professores tutores das respectivas turmas . </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>Os professores </ATOR>
<ACAO>informam aos </ACAO>
<PARAMETRO>alunos das turmas que são tutores os dias para a confirmação de matrícula . </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>as mudanças </ATOR>
<ACAO>enviadas </ACAO>
<PARAMETRO>pelos professores e confirma a inscrição dos alunos . </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>Os professores </ATOR>
<ACAO>entram </ACAO>
<PARAMETRO>em um acordo sobre horários em que gostariam de ministrar suas disicplinas . </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>a diretora </ATOR>
<ACAO>organiza </ACAO>
<PARAMETRO>uma grade de horários . </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>a diretora </ATOR>
<ACAO>encaminha </ACAO>
<PARAMETRO>a mesmo para a secretaria . </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>O horário das disicplinas </ATOR>
<ACAO>é divulgado </ACAO>
<PARAMETRO>no site do PPGI O aluno de mestrado tem que entregar o documento de inscrição em disciplinas e o termo de compromisso de orientação . </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>O orientador </ATOR>
<ACAO>assina </ACAO>
<PARAMETRO>o termo de compromisso de orientação para validar sua orientação junto com o aluno de mestrado . </PARAMETRO>

</ATIVIDADE>

<ATIVIDADE>
<ATOR>o aluno </ATOR>
<ACAO>não confirmar </ACAO>
<PARAMETRO>a inscrição em disciplinas com o professor tutor , será considerado abandono de curso por parte do aluno . </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>O aluno especial </ATOR>
<ACAO>pode escolher cursar </ACAO>
<PARAMETRO>disciplinas isoladas antes de ingressar no curso . </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ACAO>Podem ser </ACAO>
<PARAMETRO>cursadas até 2 disciplinas por semestre </PARAMETRO>
, por no máximo 2 semestres . </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ACAO>Não são </ACAO>
<PARAMETRO>todas </PARAMETRO>
as disciplinas que podem ser cursadas por alunos especiais . </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>O aluno especial </ATOR>
<ACAO>deve anexar à </ACAO>
<PARAMETRO>documentação comum para inscrição em disciplinas isoladas uma carta explicando o motivo pelo qual deseja cursar a disciplina . </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>O aluno especial </ATOR>
<ACAO>deve comparecer </ACAO>
<PARAMETRO>na secretaria para confirmar se seu pedido foi DEFERIDO ou INDEFERIDO . </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ACAO>definem </ACAO>
<PARAMETRO>as disciplinas que </PARAMETRO>
serão cursadas no semestre . </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>O aluno de mestrado </ATOR>
<ACAO>tem </ACAO>
<PARAMETRO>que escolher as disciplinas juntamente com o seu orientador . </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>A inclusão das disciplinas do aluno </ATOR>
<ACAO>é realizada </ACAO>
<PARAMETRO>no sistema após a entrega do requerimento de inscrição em disciplinas na secretaria , assinado pelo orientador e pelo próprio aluno . </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>o aluno </ATOR>
<ACAO>ainda tem </ACAO>
<PARAMETRO>um período para inclusão e exclusão de disciplinas . </PARAMETRO>

</ATIVIDADE>

<ATIVIDADE>
<ATOR>exclusão de disciplinas nos cursos do PPGI </ATOR>
<ACAO>pode ser feita </ACAO>
<PARAMETRO>num prazo aproximado de 30 dias após o fim do período regular de matrícula . </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>O aluno que </ATOR>
<ACAO>trancar </ACAO>
<PARAMETRO>uma disciplina até esse período não é considerado reprovado . </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>Se </ATOR>
<ACAO>não trancar </ACAO>
<PARAMETRO>, deve cursar até o fim , ou será considerado abandono . </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ACAO>Há </ACAO>
<PARAMETRO>um período de discussão de disciplinas a </PARAMETRO>
serem ofertadas e horários , com base em uma grade preliminar enviada pela Escola / </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>Programa . </ATOR>
<ACAO>Há </ACAO>

<PARAMETRO>um período de discussão de disciplinas a serem ofertadas e horários , com base em uma grade preliminar enviada pela Escola / </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>O professor responsável pela disciplina </ATOR>
<ACAO>deve avaliar </ACAO>
<PARAMETRO>o pedido de matrícula em disciplina isolada e encaminhar notificação de deferimento ou não do pedido . </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>O professor responsável pela disciplina </ATOR>
<ACAO>tem </ACAO>
<PARAMETRO>total autonomia para deferir ou indeferir tais pedidos . </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>as inscrições em disciplina </ATOR>
<ACAO>isolada </ACAO>
<PARAMETRO>deferidas , para posterior emissão de declaração e controle . </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>Aluno e tutor se </ATOR>
<ACAO>reúnem </ACAO>
<PARAMETRO>para verificar o histórico do aluno , suas necessidades e sugerir alterações em sua inscrição , visando o melhor aproveitamento do curso . </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>os alunos para </ATOR>
<ACAO>confirmar </ACAO>
<PARAMETRO>a matrícula nas disciplinas e encaminha as mudanças à secretária . </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>Alunos de mestrado </ATOR>
<ACAO>precisam </ACAO>
<PARAMETRO>da assinatura do orientador no requerimento . </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>O aluno também </ATOR>
<ACAO>pode </ACAO>
<PARAMETRO>, durante o período pré - determinado pela UNIRIO (</PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>PROGRAD), </ATOR>
<ACAO>trancar </ACAO>
<PARAMETRO>a matrícula no curso . </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>o trancamento </ATOR>
<ACAO>é </ACAO>
<PARAMETRO>de no máximo 1 semestre e não pode ser feito no primeiro semestre . </PARAMETRO>
</ATIVIDADE>

- Arquivo de Log de Atividades (Estudo de Caso II / Gramática I):

<ATIVIDADE>
<ATOR>O gerente de projeto </ATOR>
<ACAO>envia </ACAO>
<PARAMETRO>email para o cliente para agendamento da reunião de kick - off </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ACAO>É </ACAO>
<PARAMETRO>necessária uma reunião para </PARAMETRO>
alinhar os objetivos </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>O gerente de projeto </ATOR>
<ACAO>prepara juntamente </ACAO>
<PARAMETRO>com a equipe </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>O Escritório de Processos </ATOR>
<ACAO>envia </ACAO>
<PARAMETRO>ao cliente toda a documentação existente sobre as regras para modelagem no repositório de processos de </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>O cliente </ATOR>
<ACAO>expõe </ACAO>

<PARAMETRO>a sua necessidade </PARAMETRO>
</ATIVIDADE>

<TRIGGER>do</TRIGGER>
<ATIVIDADE>
<ATOR>Escritório de Processos </ATOR>
<ACAO>apresenta </ACAO>
<PARAMETRO>como é a gestão de um projeto de modelagem com as etapas de criação da infraestrutura </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>reunião </ATOR>
<ACAO>resulta </ACAO>
<PARAMETRO>a lista dos especialistas a serem entrevistados para levantamento das informações </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>O gerente de projeto </ATOR>
<ACAO>prepara </ACAO>
<PARAMETRO>agenda de entrevistas </PARAMETRO>
</ATIVIDADE>

<TRIGGER>do</TRIGGER>
<ATIVIDADE>
<ATOR>projeto </ATOR>
<ACAO>deverá levantar </ACAO>
<PARAMETRO>as informações para estabelecer o escopo </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>Isto </ATOR>

<ACAO>implica </ACAO>
<PARAMETRO>na definição dos macro - processos </PARAMETRO>
</ATIVIDADE>

<TRIGGER>e</TRIGGER>
<ATIVIDADE>
<ATOR>processos a </ATOR>
<ACAO>serem </ACAO>
<PARAMETRO>modelados </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>O Escritório de Processos</ATOR>
<ACAO>solicita </ACAO>
<PARAMETRO>para Administrador a liberação </PARAMETRO>
</ATIVIDADE>

<TRIGGER>do</TRIGGER>
<ATIVIDADE>
<ATOR>projeto </ATOR>
<ACAO>não pode ser iniciada </ACAO>
<PARAMETRO></PARAMETRO>
</ATIVIDADE>

<TRIGGER>caso</TRIGGER>
<ATIVIDADE>
<ATOR>o Administrador </ATOR>
<ACAO>julgue </ACAO>
<PARAMETRO></PARAMETRO>
</ATIVIDADE>

<TRIGGER>que</TRIGGER>
<ATIVIDADE>

<ATOR>o mesmo </ATOR>
<ACAO>não está </ACAO>
<PARAMETRO>alinhado à iniciativa estratégica de negócio </PARAMETRO>
</ATIVIDADE>

<TRIGGER>E</TRIGGER>
<ATIVIDADE>
<ATOR>Área. Administrador </ATOR>
<ACAO>autoriza </ACAO>
<PARAMETRO>a modelagem de processos no repositório corporativo de </PARAMETRO>
</ATIVIDADE>

<TRIGGER>E</TRIGGER>
<ATIVIDADE>
<ATOR>a equipe de modelagem </ATOR>
<ACAO>iniciar </ACAO>
<PARAMETRO>o projeto </PARAMETRO>
</ATIVIDADE>

<TRIGGER>do</TRIGGER>
<ATIVIDADE>
<ATOR>os processos </ATOR>
<ACAO>seguindo </ACAO>
<PARAMETRO>os padrões de modelagem de processos </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>O Escritório de Processos </ATOR>
<ACAO>realiza </ACAO>
<PARAMETRO>a cada 15 dias uma auditoria básica no projeto que está sendo modelado </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>São </ATOR>
<ACAO>executados </ACAO>
<PARAMETRO>os scripts do ARIS com categoria Auditoria </PARAMETRO>
</ATIVIDADE>

- Arquivo de Log de Atividades (Estudo de Caso II / Gramática II):

<ATIVIDADE>
<ATOR>O gerente de projeto </ATOR>
<ACAO>envia </ACAO>
<PARAMETRO>email para o cliente para agendamento da reunião de kick - off . </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ACAO>É </ACAO>
<PARAMETRO>necessária uma reunião para </PARAMETRO>
alinhar os objetivos do projeto . </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>O gerente de projeto </ATOR>
<ACAO>prepara juntamente </ACAO>
<PARAMETRO>com a equipe do Escritório de Processos os documentos para a reunião . </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>O Escritório de Processos </ATOR>
<ACAO>envia </ACAO>
<PARAMETRO>ao cliente toda a documentação existente sobre as regras para modelagem no repositório de processos da Área . São elas : Instruções para Modelagem de Processos , Diretrizes de Modelagem de Processos da Área . O cliente expõe a sua necessidade . </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>A equipe do Escritório de Processos </ATOR>
<ACAO>apresenta </ACAO>
<PARAMETRO>como é a gestão de um projeto de modelagem com as etapas de criação da infraestrutura (</PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>reunião </ATOR>
<ACAO>resulta </ACAO>
<PARAMETRO>a lista dos especialistas a serem entrevistados para levantamento das informações . </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>O gerente de projeto </ATOR>
<ACAO>prepara </ACAO>
<PARAMETRO>agenda de entrevistas e a envia para o cliente . </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>Nesta etapa a equipe do projeto </ATOR>
<ACAO>deverá levantar </ACAO>
<PARAMETRO>as informações para estabelecer o escopo e objetivos do projeto . </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>Isto </ATOR>
<ACAO>implica </ACAO>
<PARAMETRO>na definição dos macro - processos e processos a serem modelados . </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR> O Escritório de Processos </ATOR>
<ACAO>solicita </ACAO>
<PARAMETRO>para Administrador a liberação do projeto . </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>A modelagem do projeto </ATOR>
<ACAO>não pode ser iniciada </ACAO>
<PARAMETRO>caso o Administrador julgue que o mesmo não está alinhado à iniciativa estratégica de negócio da Área . Administrador autoriza a modelagem de processos no repositório corporativo da Área . O Escritório de Processos disponibliza os ambientes necessários para a equipe de modelagem iniciar o projeto . </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>os processos </ATOR>
<ACAO>seguindo </ACAO>
<PARAMETRO>os padrões de modelagem de processos e de estruturas de grupos de diretórios do repositório ARIS . </PARAMETRO>
</ATIVIDADE>

<ATIVIDADE>
<ATOR>O Escritório de Processos </ATOR>
<ACAO>realiza </ACAO>
<PARAMETRO>a cada 15 dias uma auditoria básica no projeto que está sendo modelado . </PARAMETRO>
</ATIVIDADE>

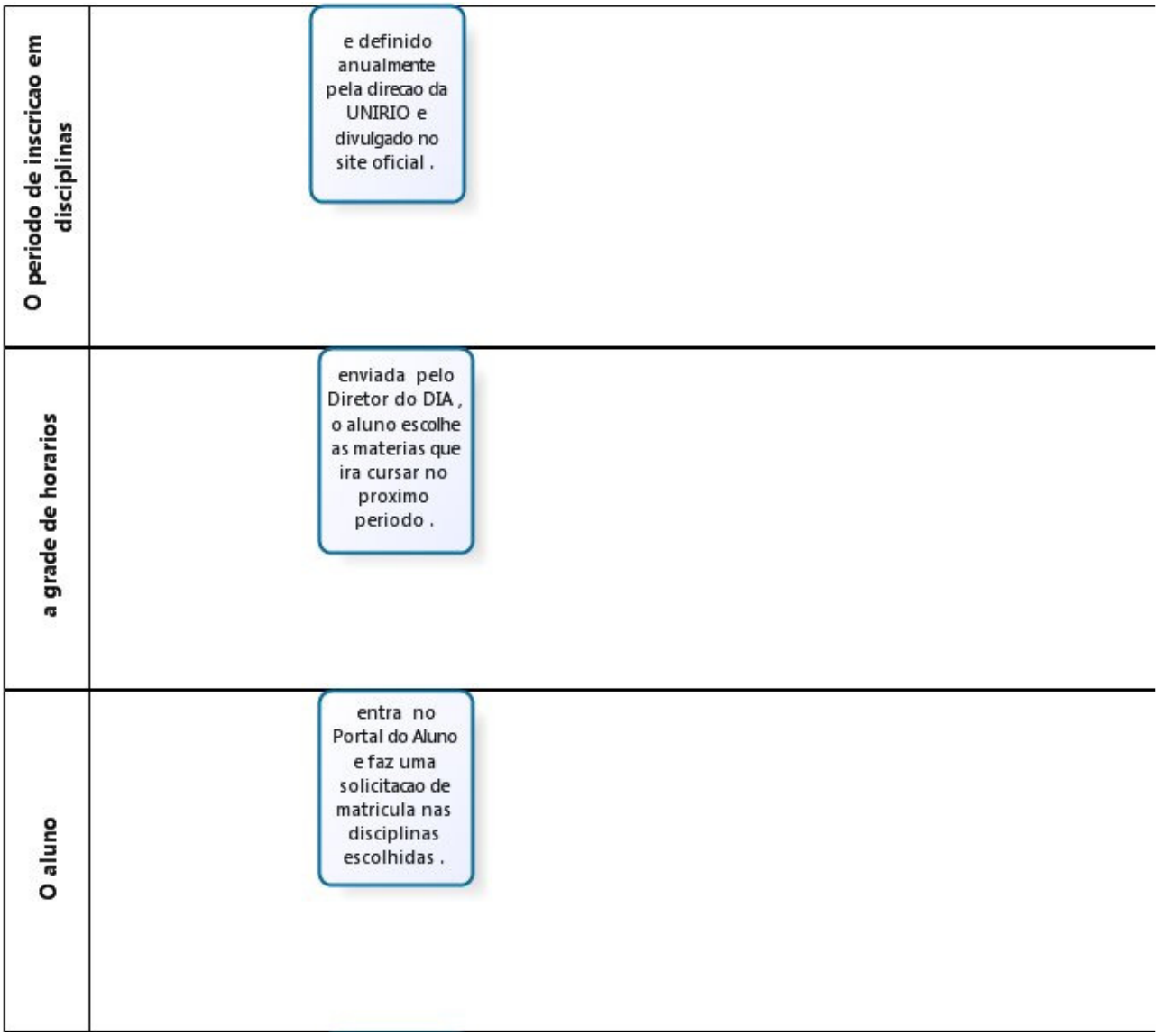
<ATIVIDADE>
<ATOR>São </ATOR>

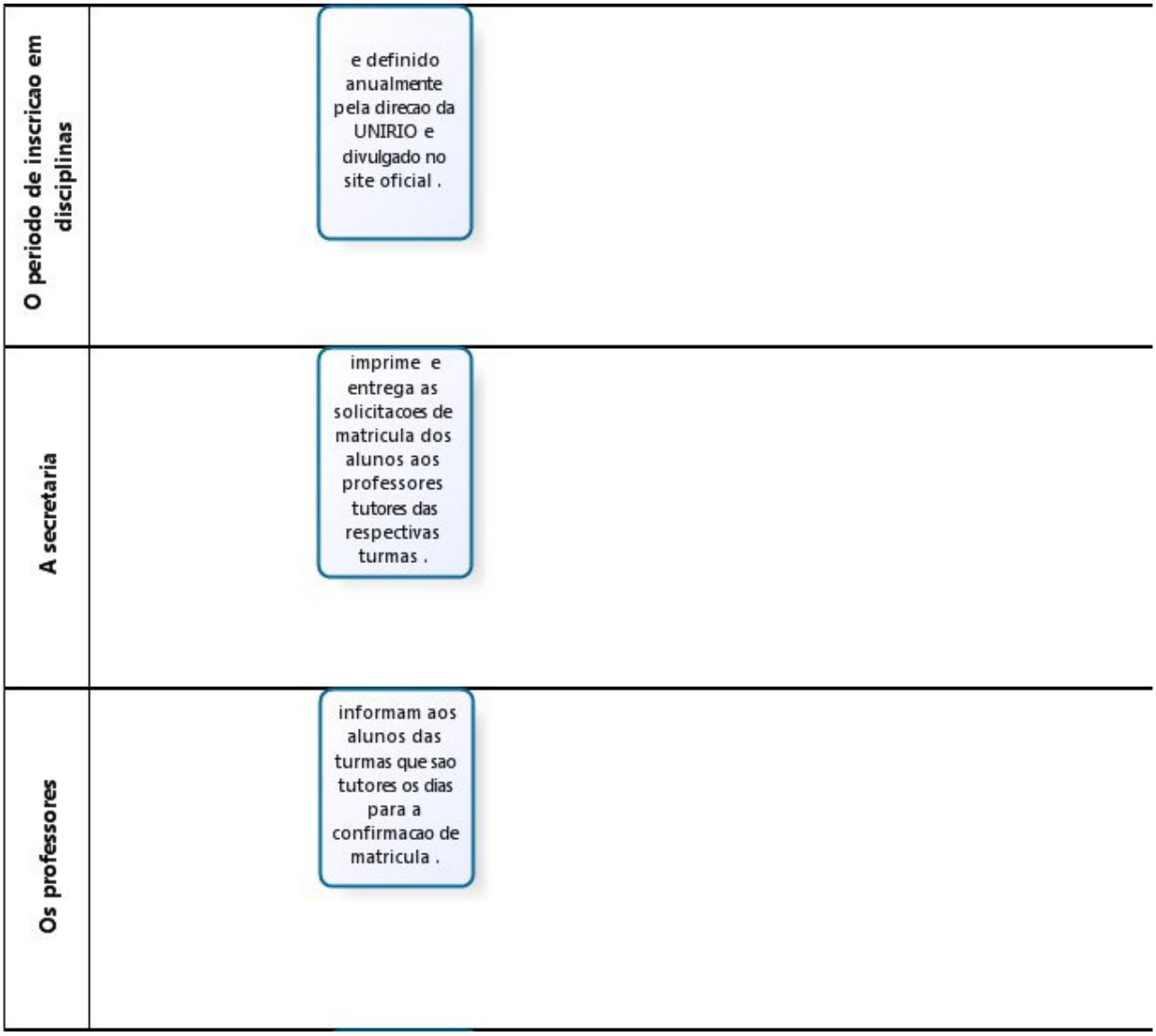
<ACAO>executados </ACAO>

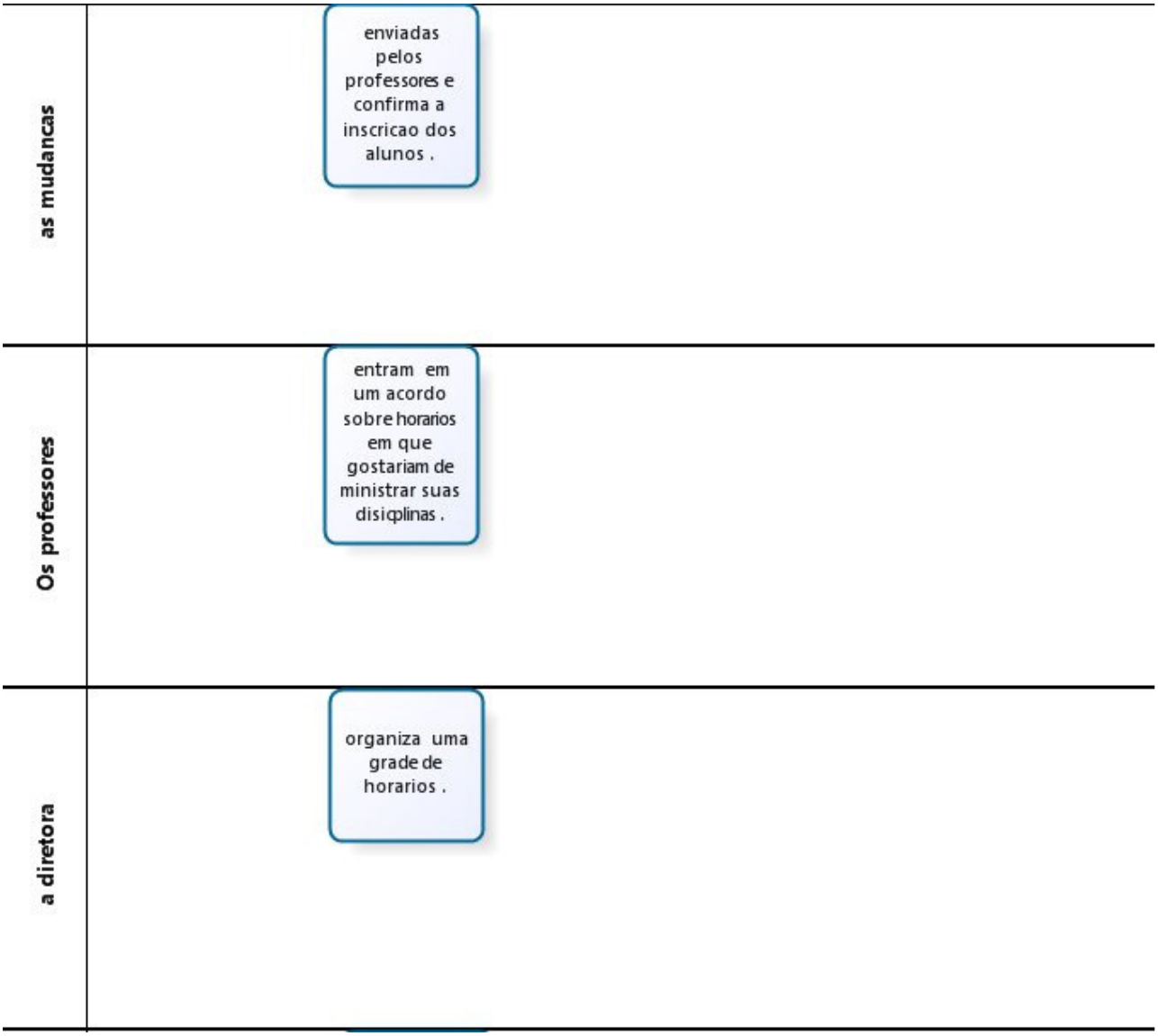
<PARAMETRO>os scripts do ARIS com categoria Auditoria e reportado ao cliente inconformidades com os padrões de modelagem e de estrutura de grupos . </PARAMETRO>

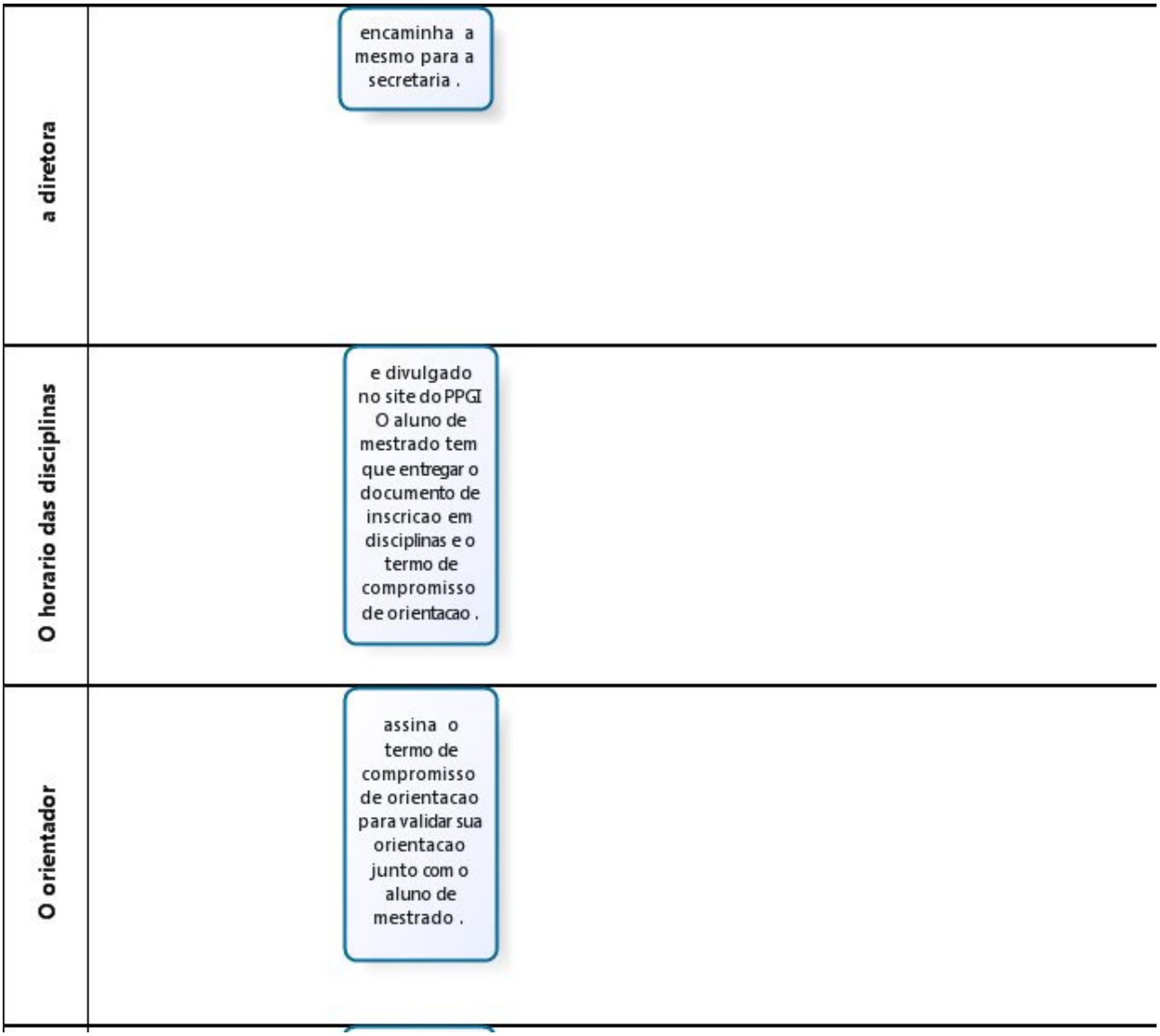
</ATIVIDADE>

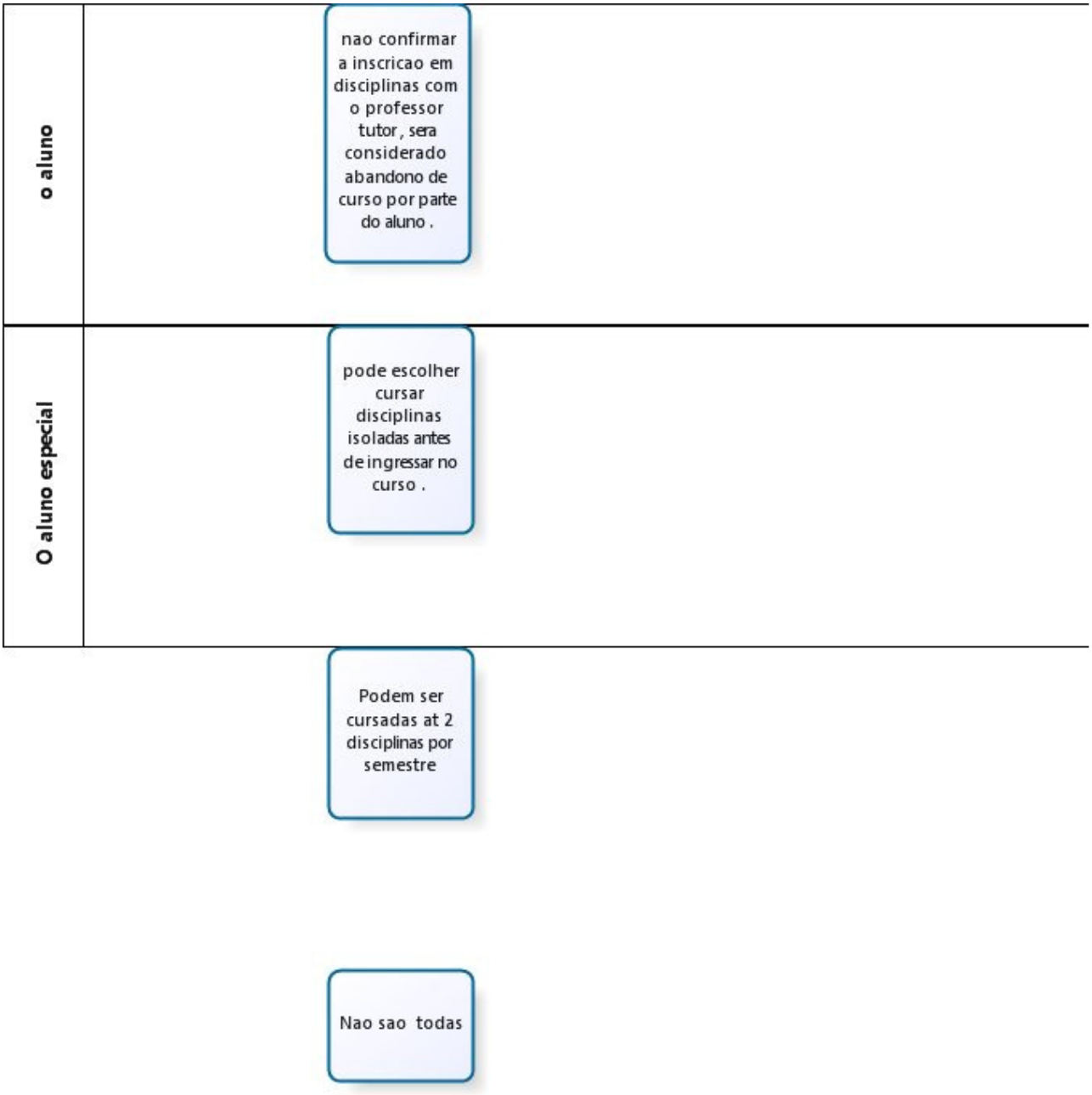
- Modelo BPMN (Estudo de Caso I / Gramática II):

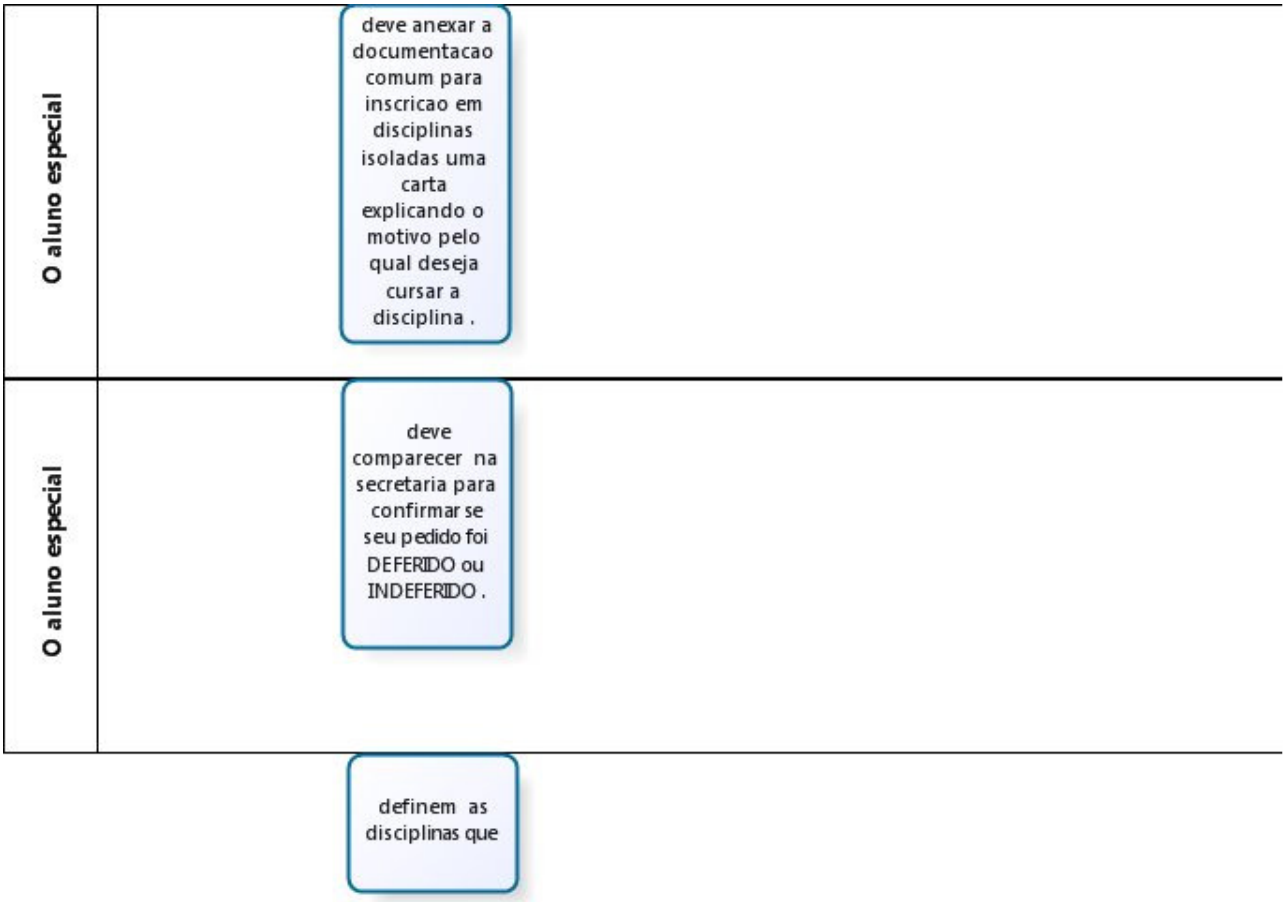


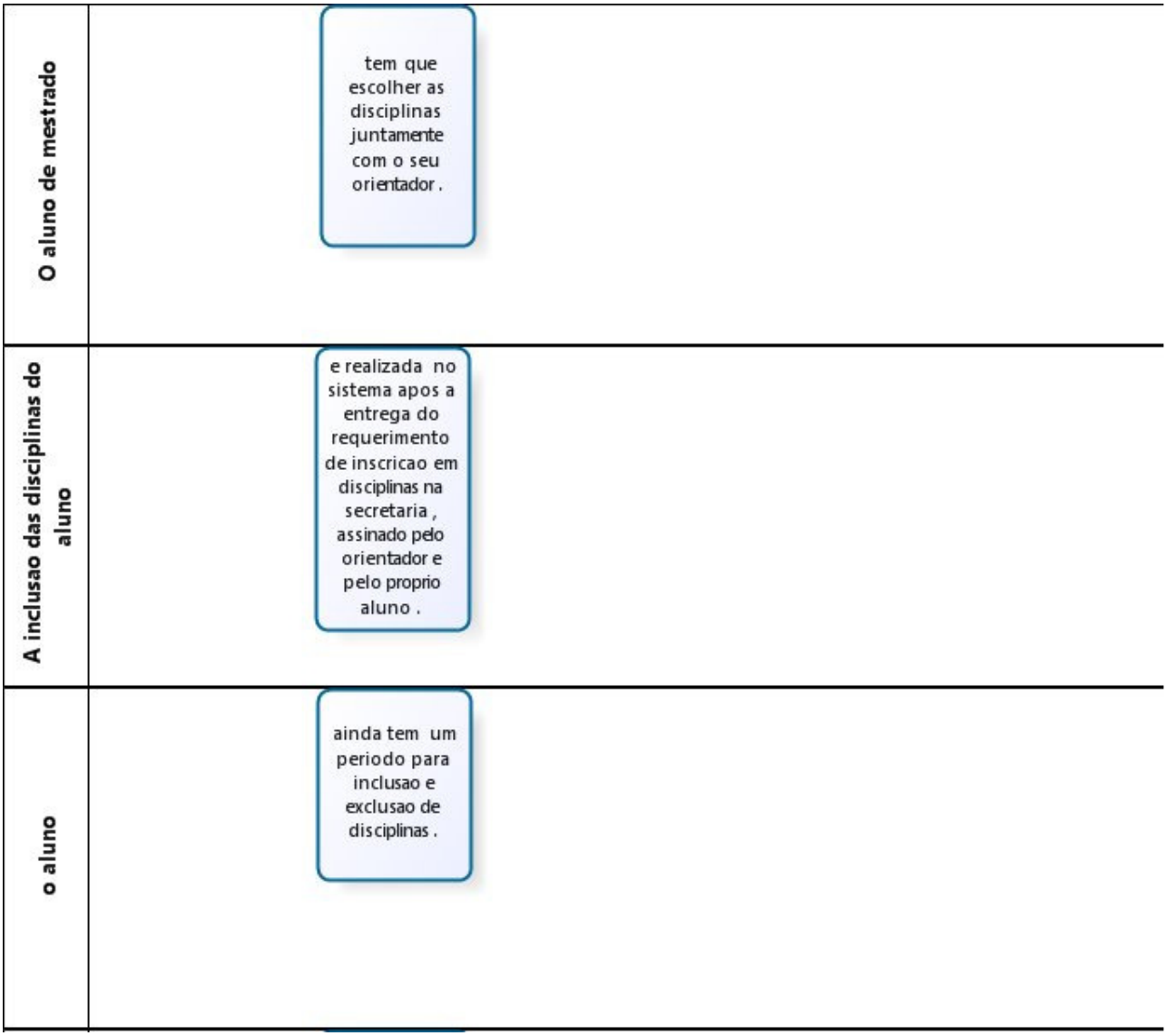

O periodo de inscricao em disciplinas
e definido anualmente pela direcao da UNIRIO e divulgado no site oficial .
A secretaria
imprime e entrega as solicitacoes de matricula dos alunos aos professores tutores das respectivas turmas .
Os professores
informam aos alunos das turmas que sao tutores os dias para a confirmacao de matricula .

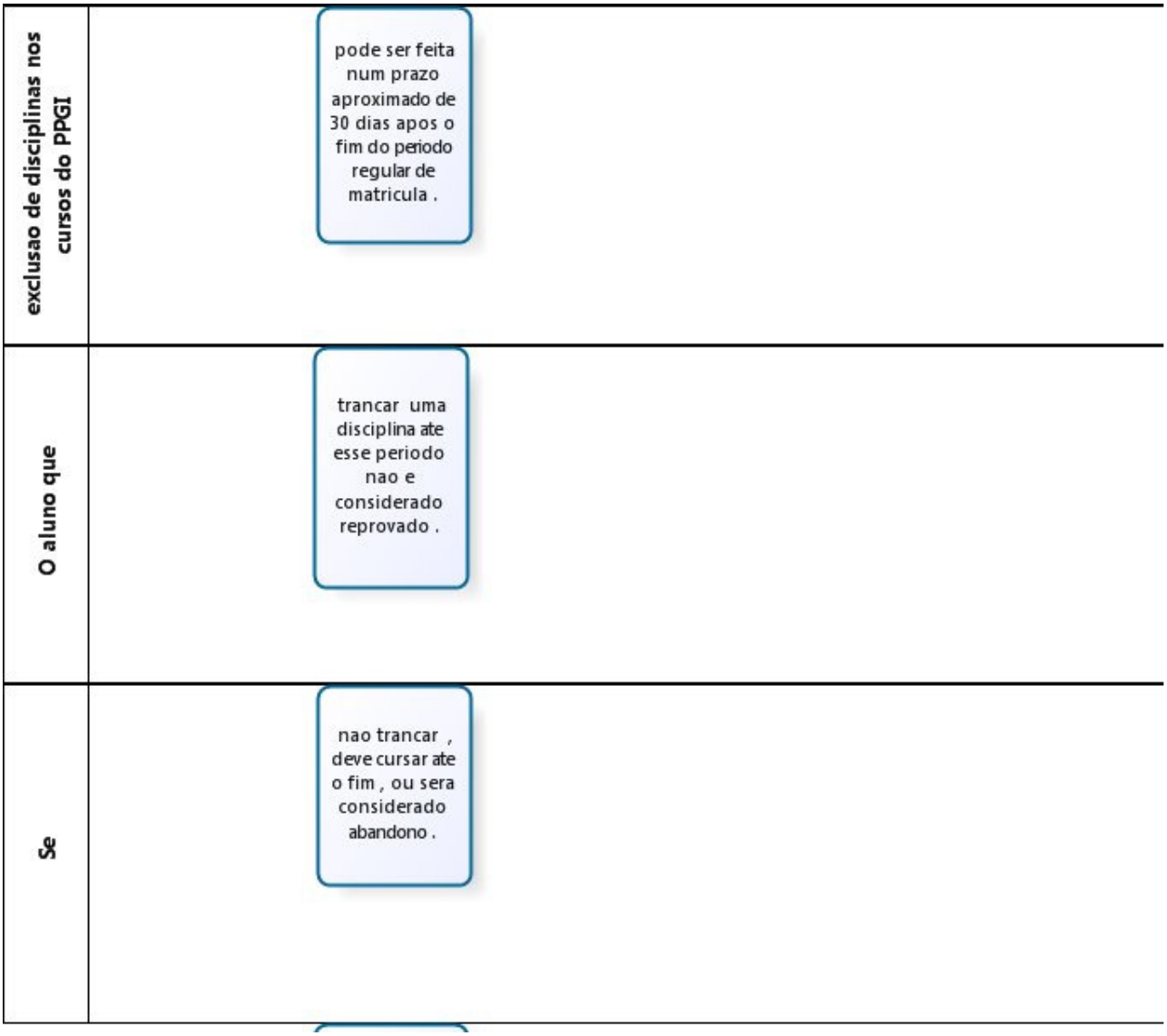

as mudancas
enviadas pelos professores e confirma a inscricao dos alunos .
Os professores
entram em um acordo sobre horarios em que gostariam de ministrar suas disiqplinas .
a diretora
organiza uma grade de horarios .

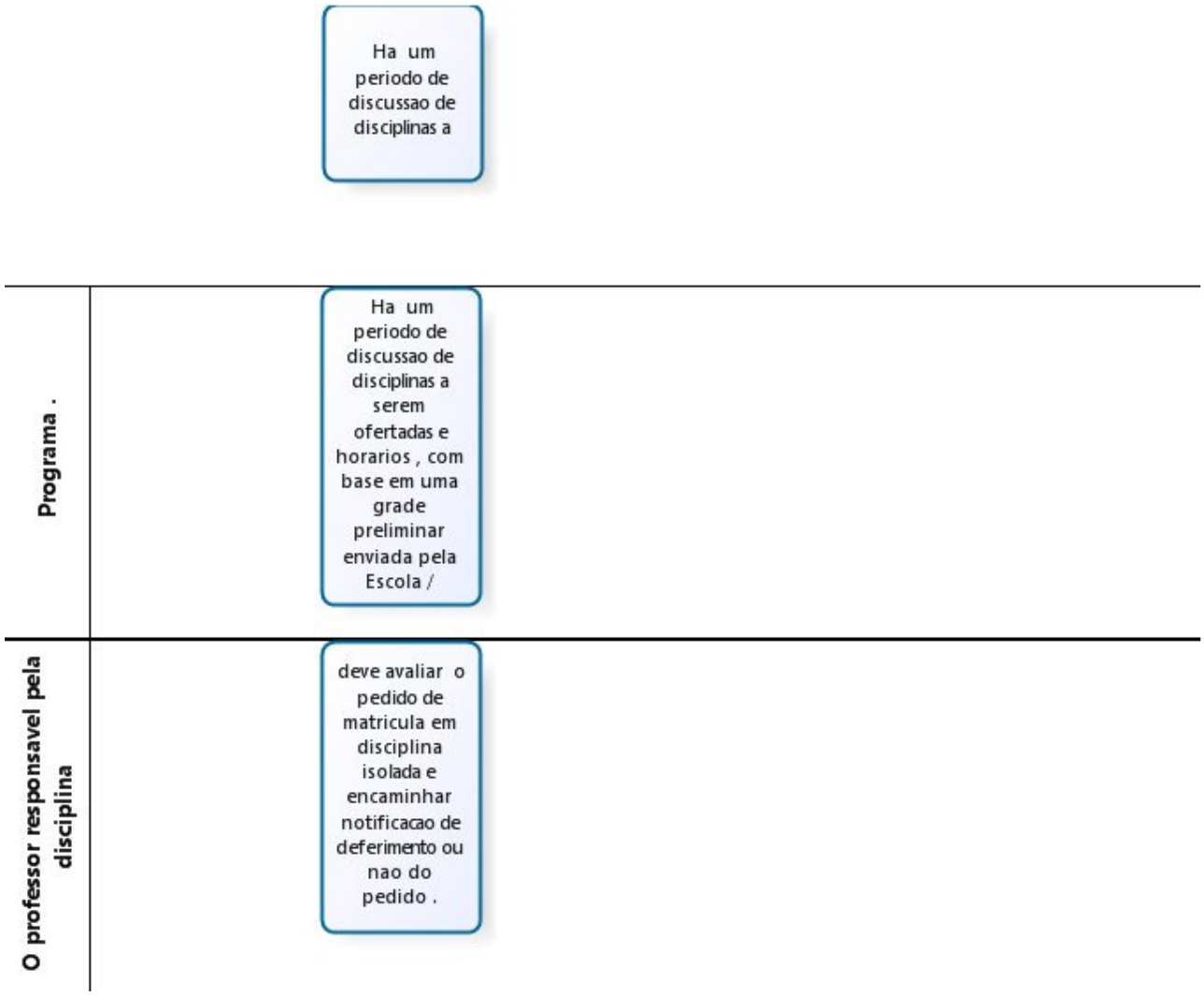

a diretora
encaminha a mesmo para a secretaria .
O horario das disciplinas
e divulgado no site do PPGI O aluno de mestrado tem que entregar o documento de inscricao em disciplinas e o termo de compromisso de orientacao .
O orientador
assina o termo de compromisso de orientacao para validar sua orientacao junto com o aluno de mestrado .

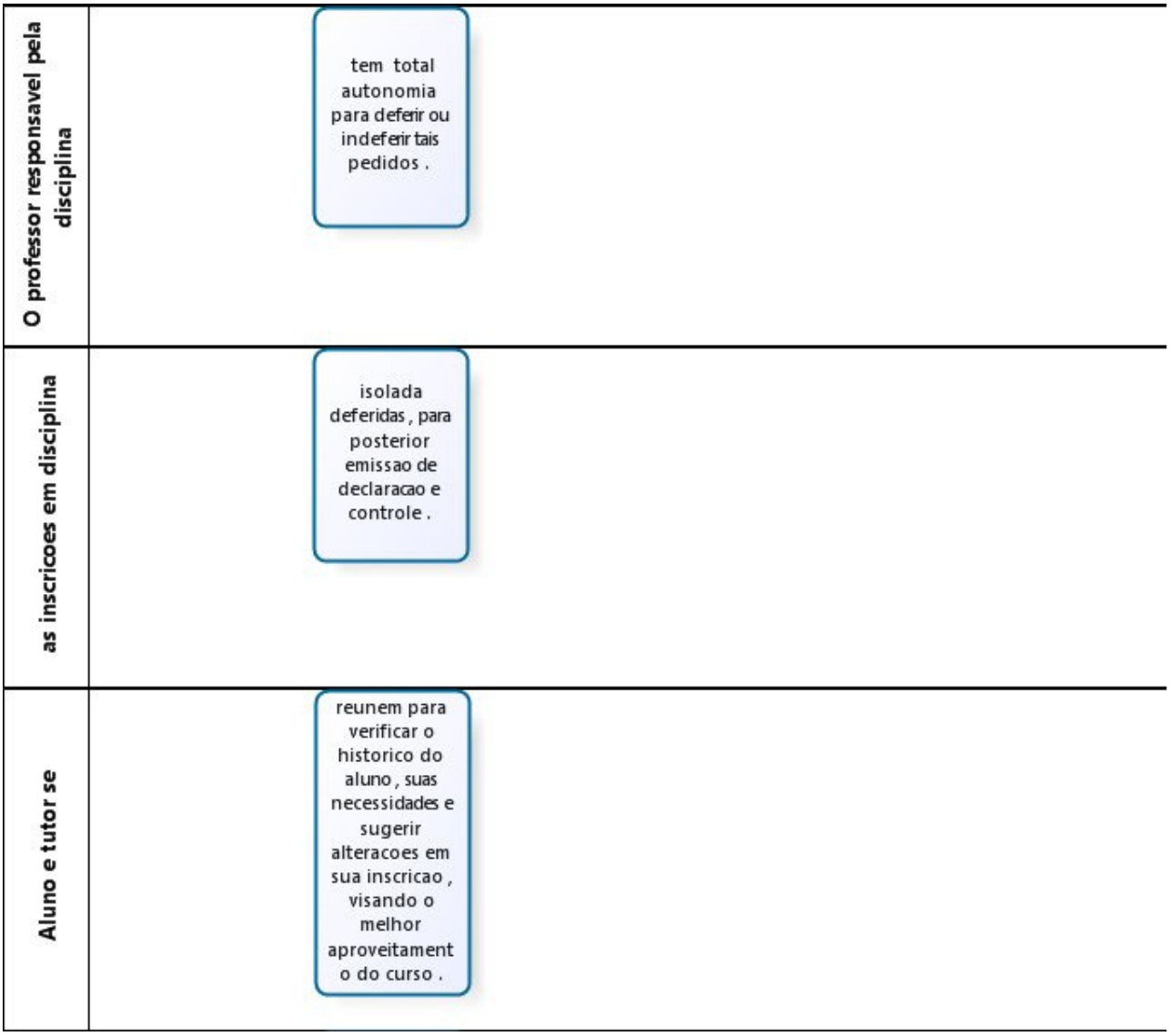

o aluno
nao confirmar a inscricao em disciplinas com o professor tutor, sera considerado abandono de curso por parte do aluno .
O aluno especial
pode escolher cursar disciplinas isoladas antes de ingressar no curso .
Podem ser cursadas at 2 disciplinas por semestre
Nao sao todas

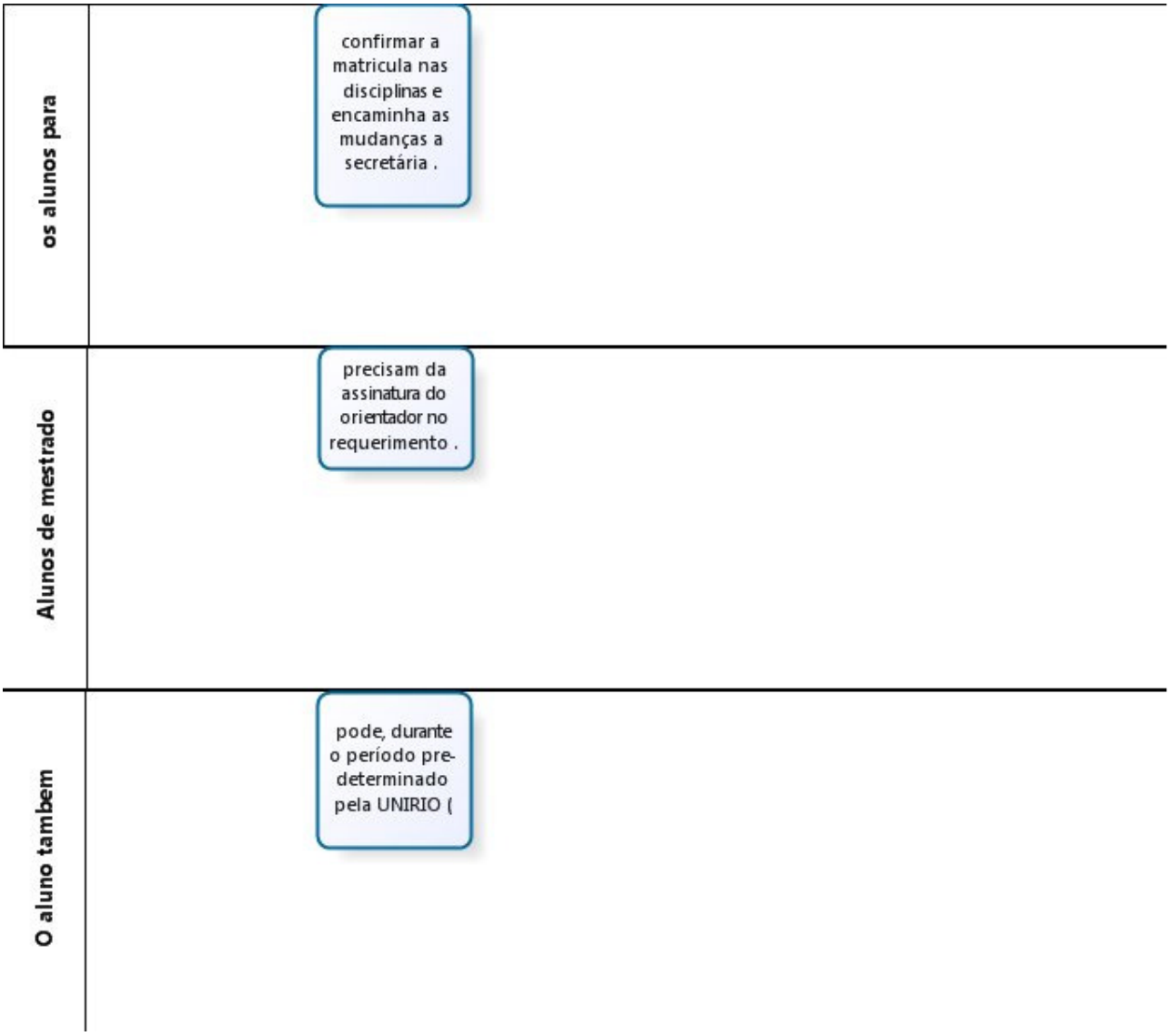

O aluno especial
deve anexar a documentacao comum para inscricao em disciplinas isoladas uma carta explicando o motivo pelo qual deseja cursar a disciplina .
O aluno especial
deve comparecer na secretaria para confirmar se seu pedido foi DEFERIDO ou INDEFERIDO .
definem as disciplinas que

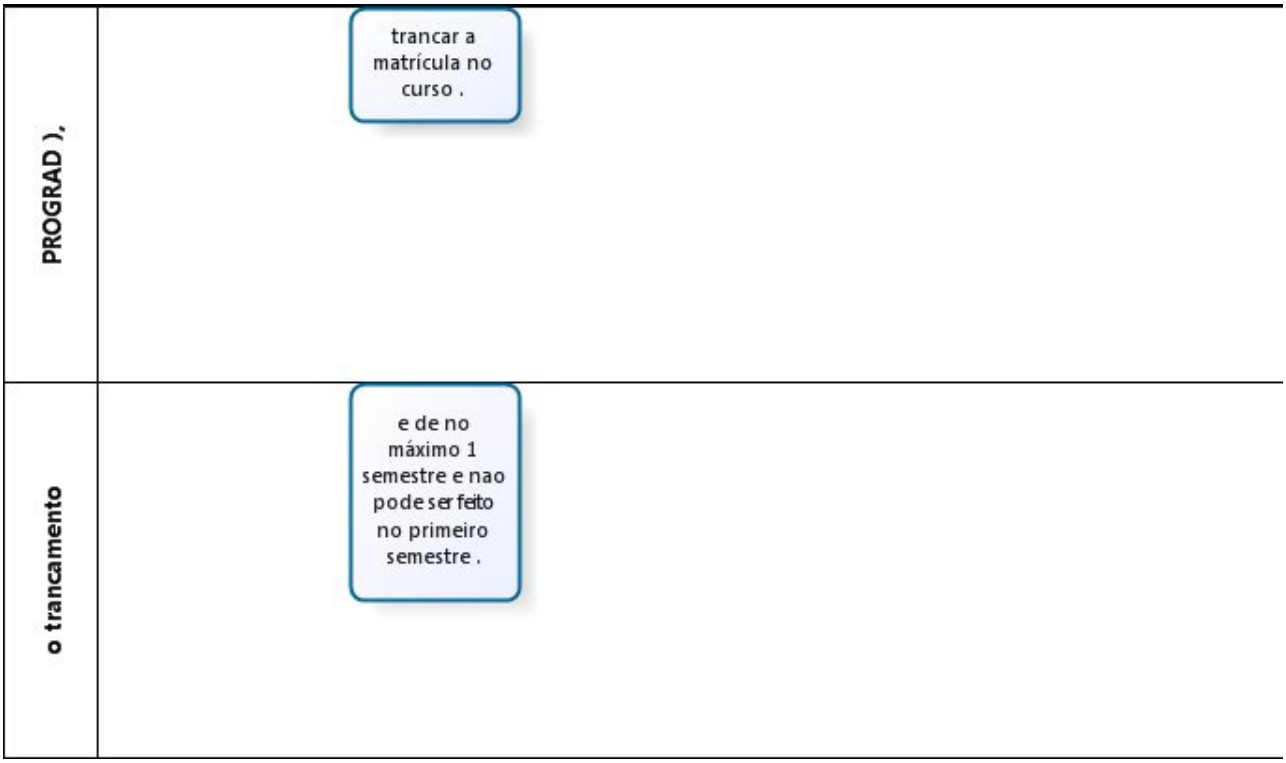

O aluno de mestrado
tem que escolher as disciplinas juntamente com o seu orientador .
A inclusao das disciplinas do aluno
e realizada no sistema apos a entrega do requerimento de inscricao em disciplinas na secretaria , assinado pelo orientador e pelo proprio aluno .
o aluno
ainda tem um periodo para inclusao e exclusao de disciplinas .


exclusao de disciplinas nos cursos do PPGI
pode ser feita num prazo aproximado de 30 dias apos o fim do periodo regular de matricula .
O aluno que
trancar uma disciplina ate esse periodo nao e considerado reprovado .
Se
nao trancar , deve cursar ate o fim , ou sera considerado abandono .

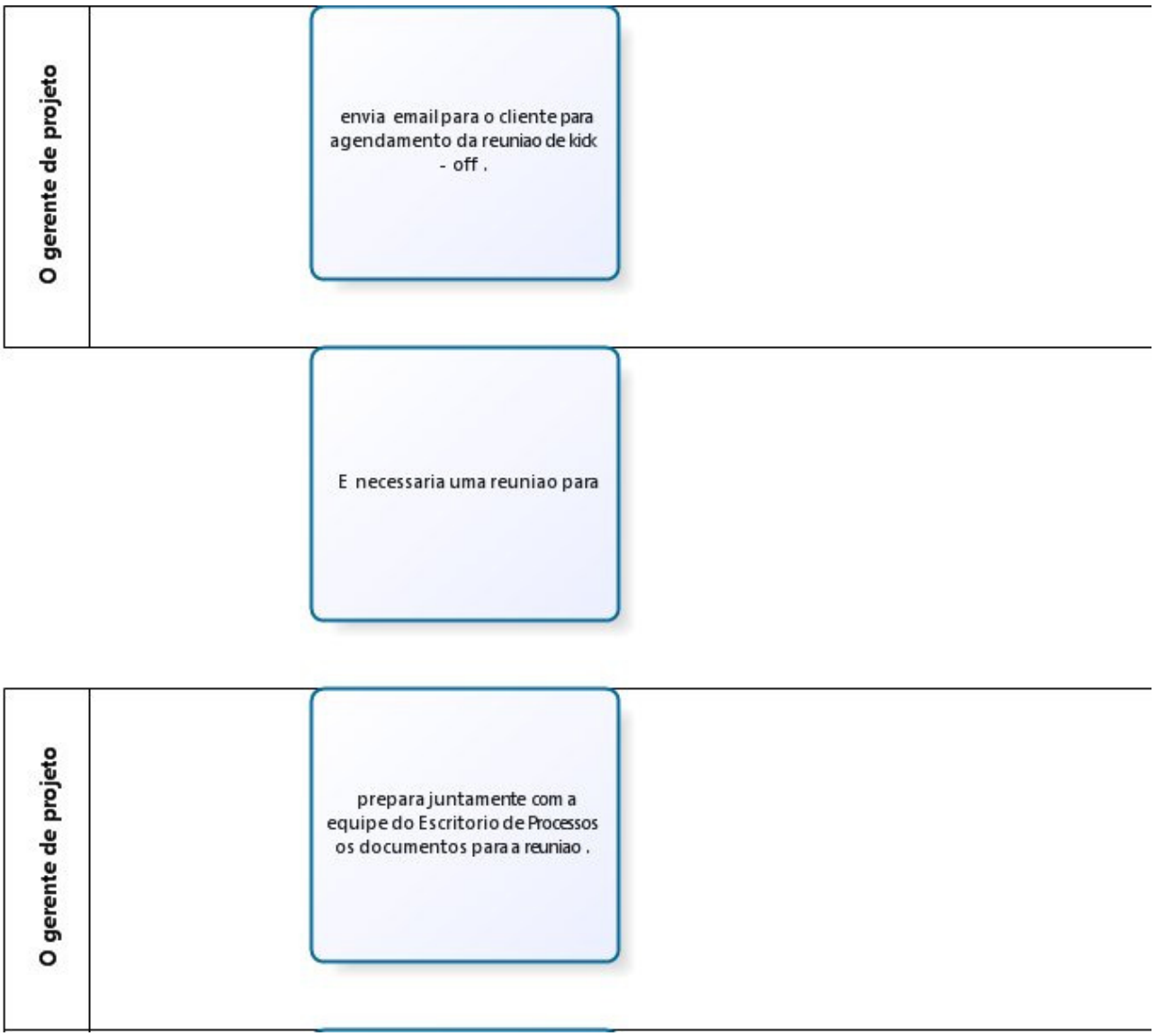

Ha um periodo de discussao de disciplinas a
Programa .
Ha um periodo de discussao de disciplinas a serem ofertadas e horarios , com base em uma grade preliminar enviada pela Escola /
O professor responsavel pela disciplina
deve avaliar o pedido de matricula em disciplina isolada e encaminhar notificacao de deferimento ou nao do pedido .

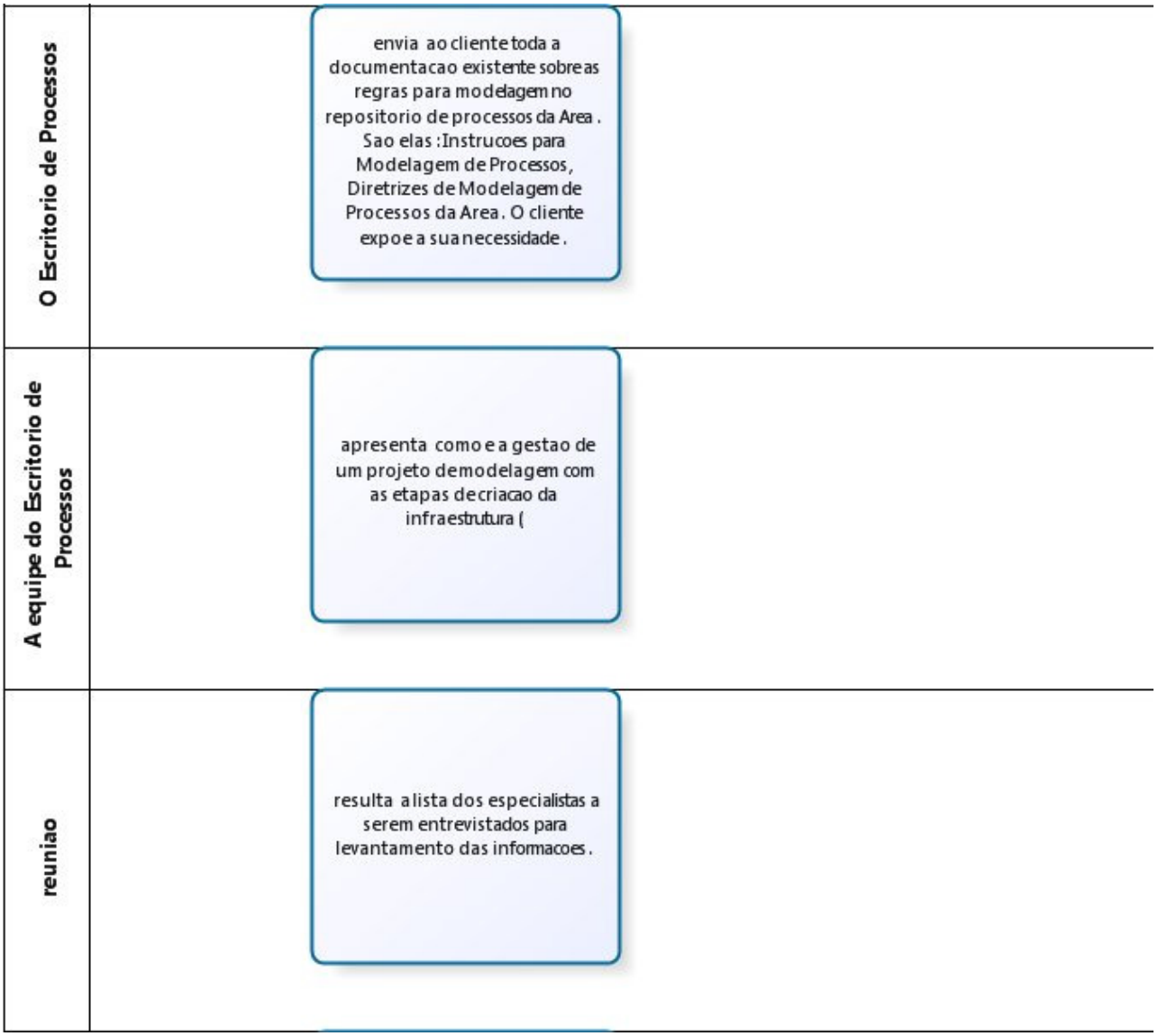

O professor responsavel pela disciplina
tem total autonomia para deferir ou indeferir tais pedidos .
as inscricoes em disciplina
isolada deferidas , para posterior emissao de declaracao e controle .
Aluno e tutor se
reunem para verificar o historico do aluno , suas necessidades e sugerir alteracoes em sua inscricao , visando o melhor aproveitament o do curso .


os alunos para
confirmar a matricula nas disciplinas e encaminha as mudanças a secretária .
Alunos de mestrado
precisam da assinatura do orientador no requerimento .
O aluno tambem
pode, durante o período pre-determinado pela UNIRIO (

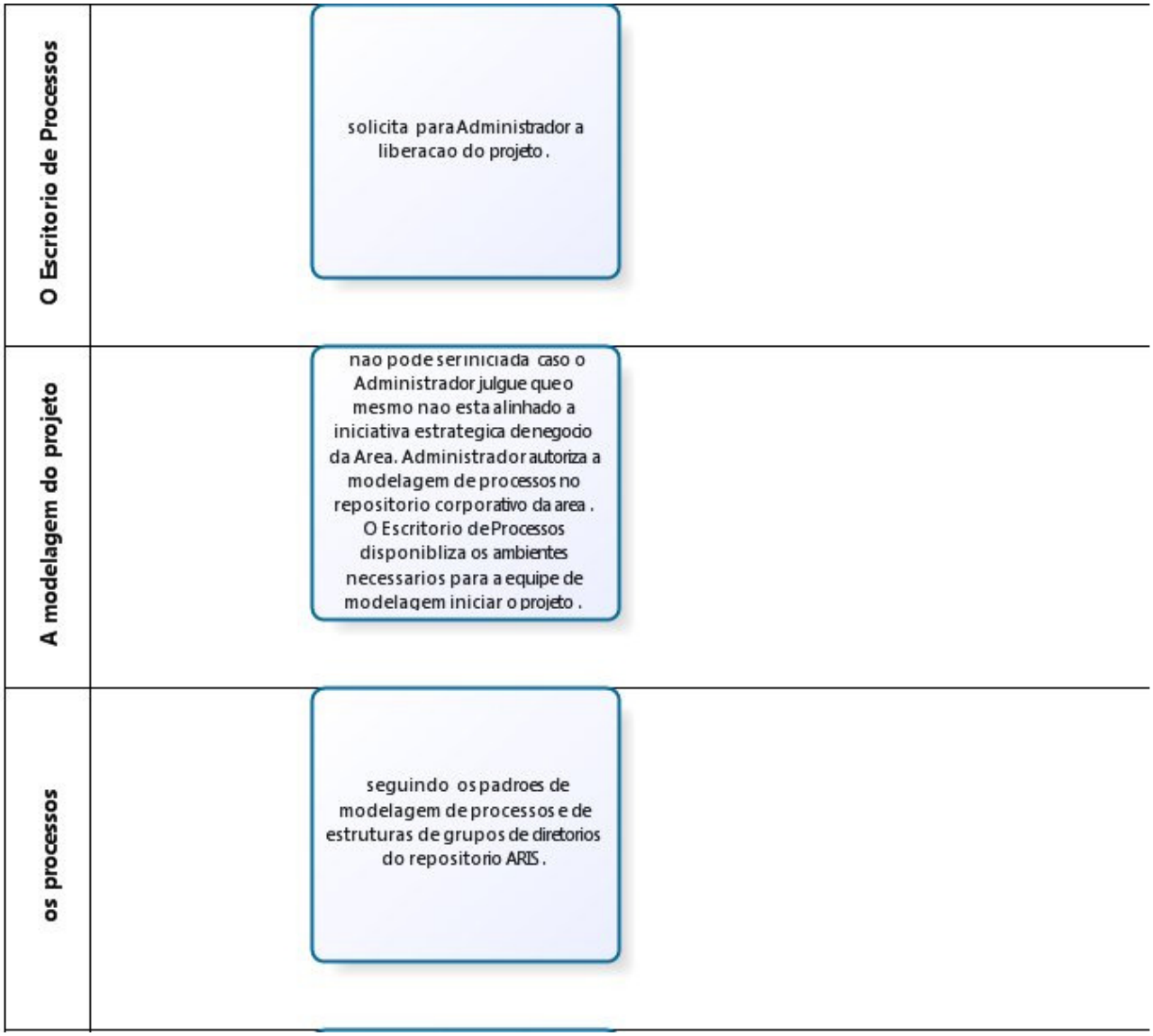

PROGRAD),
trancar a matrícula no curso .
o trancamento
e de no máximo 1 semestre e nao pode ser feito no primeiro semestre .

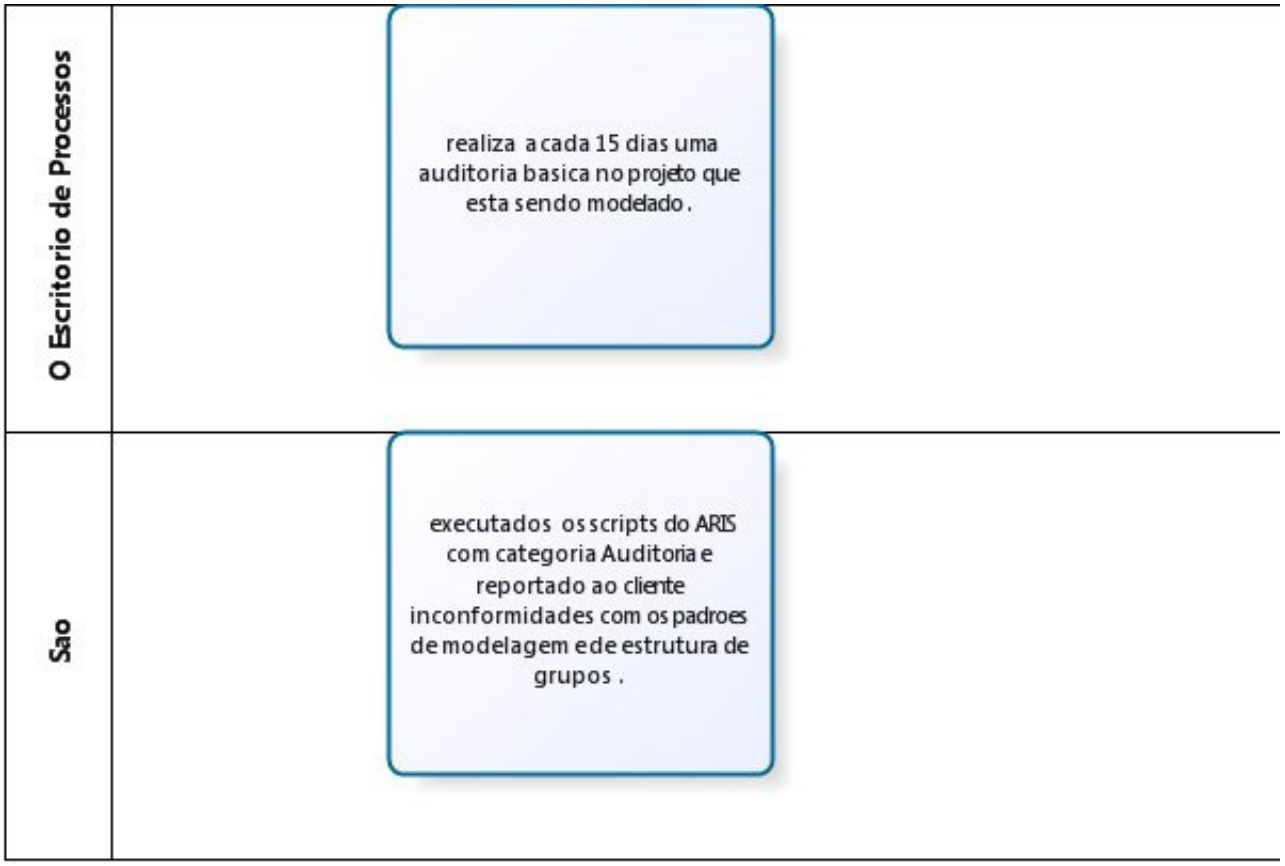
- Modelo BPMN (Estudo de Caso II / Gramática II):




O Escritorio de Processos
envia ao cliente toda a documentacao existente sobre as regras para modelagem no repositorio de processos da Area . Sao elas :Instrucoes para Modelagem de Processos, Diretrizes de Modelagem de Processos da Area . O cliente expoe a sua necessidade .
A equipe do Escritorio de Processos
apresenta como e a gestao de um projeto de modelagem com as etapas de criacao da infraestrutura (
reuniao
resulta a lista dos especialistas a serem entrevistados para levantamento das informacoes .

O gerente de projeto
prepara agenda de entrevistas e a envia para o cliente .
Nesta etapa a equipe do projeto
devera levantar as informacoes para estabelecer o escopo e objetivos do projeto .
Isto
implica na definicao dos macro - processos e processos a serem modelados .


O Escritorio de Processos
solicita para Administrador a liberacao do projeto.
A modelagem do projeto
nao pode ser iniciada caso o Administrador julgue que o mesmo nao esta alinhado a iniciativa estrategica de negocio da Area. Administrador autoriza a modelagem de processos no repositorio corporativo da area. O Escritorio de Processos disponibliza os ambientes necessarios para a equipe de modelagem iniciar o projeto.
os processos
seguindo os padroes de modelagem de processos e de estruturas de grupos de diretorios do repositorio ARIS.


O Escritorio de Processos
realiza a cada 15 dias uma auditoria basica no projeto que esta sendo modelado .
Sao
executados os scripts do ARIS com categoria Auditoria e reportado ao cliente inconformidades com os padroes de modelagem e de estrutura de grupos .