



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

UM MÉTODO PARA DESCOBERTA DE PROCESSOS INTENSIVOS
EM CONHECIMENTO A PARTIR DA MINERAÇÃO DE E-MAILS

Diego Carvalho Soares

Orientadora

Flávia Maria Santoro

Coorientadora

Fernanda Araujo Baião Amorim

RIO DE JANEIRO, RJ – BRASIL

SETEMBRO DE 2011

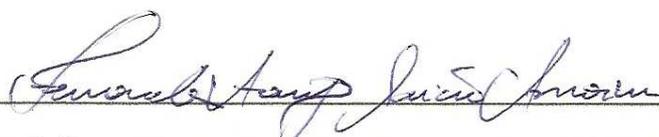
UM MÉTODO PARA DESCOBERTA DE PROCESSOS INTENSIVOS EM
CONHECIMENTO A PARTIR DA MINERAÇÃO DE E-MAILS

Diego Carvalho Soares

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.



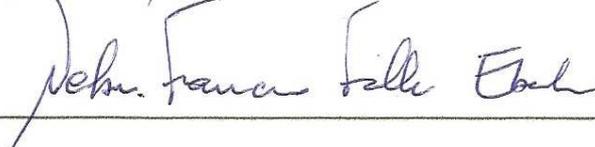
Prof^a. Flávia Maria Santoro, D.Sc. - UNIRIO



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



Prof^a. Renata Mendes de Araujo, D.Sc. - UNIRIO



Prof. Nelson Francisco Favilla Ebecken, Dr. – COPPE/UFRJ

RIO DE JANEIRO, RJ – BRASIL

SETEMBRO DE 2011

S676 Soares, Diego Carvalho.
Um Método para Descoberta de Processos Intensivos em Conhecimento a partir da Mineração de E-mails / Diego Carvalho Soares, 2011.
122f. + 1 CD-ROM.

Orientadora: Flávia Maria Santoro.
Coorientadora: Fernanda Araujo Baião Amorim.
Dissertação (Mestrado em Informática) – Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro, 2011.

1. Processo Intensivo em Conhecimento. 2. Sistema de Informação. 3. Tecnologia da 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. Centro de Ciências Exatas e Tecnologia. Curso de Especialização em Informática. IV. Título.

CDD – 005.5

*À minha avó, minha mãe e minha esposa,
as mulheres da minha vida.*

“A persistência é o menor caminho do êxito”.

Charles Chaplin

Agradecimentos

Agradeço, primeiramente e acima de tudo, a Deus, por prover toda a força, iluminação, motivação e inspiração necessárias para a realização de mais esta conquista.

À minha avó e minha mãe, por estarem sempre ao meu lado, renunciando a tantas coisas em prol do meu bem estar e do meu crescimento pessoal e profissional; pelo apoio e motivação de todas as horas; por investirem e acreditarem em mim.

À minha esposa, por todo o carinho, amor, dedicação, motivação e paciência; pelas palavras de conforto; por abdicar, por diversas vezes, de momentos livres, para estar ao meu lado, provendo o apoio necessário.

Às minhas orientadoras e amigas, Flávia Santoro e Fernanda Baião, pelos conselhos certos nas horas certas; por atenderem prontamente a todas as minhas dúvidas; pelas palavras de incentivo e motivação; pelos almoços agradáveis, servidos de enxurradas de ideias; por toda a competência, seriedade e comprometimento.

A todo corpo docente do PPGI da UNIRIO, em especial a Renata Araújo, Asterio Tanaka e Márcio Barros, por toda dedicação e idoneidade com que conduzem o programa; pela oportunidade de integrar o corpo docente do PPGI; pelos bons momentos vividos nestes mais de oito anos de convivência.

Ao amigo do PPGI, João Gonçalves, por todo o apoio com o método *Story Mining*; pela disponibilidade em atender às minhas solicitações; pelos conselhos positivos e motivacionais que sempre fizeram a diferença.

A todos os colegas do PPGI pelos momentos agradáveis que passamos juntos; pelo apoio mútuo em diversas apresentações; pelas constantes trocas de conhecimento.

À minha família e amigos, pela compreensão das muitas ausências; pela constante torcida e apoio; por tudo que representam na minha vida.

Soares, Diego Carvalho. Um Método para Descoberta de Processos Intensivos em Conhecimento a partir da Mineração de E-mails. UNIRIO, 2011. 122 páginas. Dissertação de Mestrado. Departamento de Informática Aplicada, UNIRIO.

RESUMO

Conhecimento consiste em toda informação retida na mente de um indivíduo, personalizada ou relacionada a fatos, conceitos, interpretações, ideias, observações e julgamentos. A Gestão do Conhecimento apresenta-se como um meio sistemático de apoiar a identificação, desenvolvimento e provisão do conhecimento relevante ao sucesso organizacional. A Gestão do Conhecimento Orientada a Processos de Negócio defende a descoberta de conhecimento organizacional através da execução de Processos de Negócio. Neste contexto, os Processos de Negócio Intensivos em Conhecimento são considerados peça-chave, uma vez que a forte intensidade de conhecimento nestes contida torna-os fonte essencial de conhecimento organizacional. Portanto, identificar, modelar, analisar e aperfeiçoar Processos Intensivos em Conhecimento deve ser o principal objetivo da Gestão do Conhecimento Orientada a Processos. No entanto, a intensa socialização e troca informal de conhecimento presentes nestes processos fazem com que o apoio da Tecnologia da Informação (TI) aos Processos Intensivos em Conhecimento ainda não seja adequado.

Esta pesquisa tem como objetivo propor um método capaz de minerar o conteúdo textual proveniente de uma Ferramenta Colaborativa, os E-mails, de forma que seja possível extrair e representar os elementos que compõem os Processos Intensivos em Conhecimento, tornando explícito um conhecimento essencialmente tácito e, conseqüentemente, apoiando o avanço da TI na Gestão do Conhecimento organizacional.

Palavras-chave: Gestão do Conhecimento Orientada a Processos de Negócio, Processos Intensivos em Conhecimento, Ferramentas Colaborativas, Mineração de Texto.

ABSTRACT

Knowledge is a fluid mix of framed experience, contextual information, values and expert insight that provides a framework for evaluating and incorporating new experiences and information. Knowledge Management is the broad process to acquire, process and create accessible knowledge in a more systematic way, in order to obtain better decisions and to be prepared for the future. Business-Process Oriented Knowledge Management (BPOKM) considers business processes as initial point for knowledge management. Within process-oriented knowledge management the knowledge-intensive business process is the primary perspective. A knowledge-intensive process should be a core process of the company and it should produce or add new knowledge to the organization's knowledge base. However, the IT-support for knowledge-intensive business processes is generally not very sophisticated because it strongly relies on socialization and informal exchange of knowledge

This research aims to provide a new method capable of mining data generated by a collaborative tool, e-mails, in order to extract and represent knowledge-intensive processes main elements, making tacit knowledge explicit and thus supporting the advance of IT in knowledge management research area.

Keywords: Business Process Oriented Knowledge Management, Knowledge Intensive Processes, Collaboration Tools, Text Mining.

Sumário

Capítulo 1 – Introdução	1
1.1 Relevância	1
1.2 Problema.....	2
1.3 Hipótese	2
1.4 Justificativa.....	3
1.5 Objetivos	4
1.6 Delimitações de Abrangência	4
1.7 Procedimento Metodológico	5
1.8 Organização Estrutural.....	6
Capítulo 2 – Fundamentação Teórica	7
2.1 Conhecimento	7
2.1.1 Conhecimento e os Sistemas de Informação	8
2.1.2 Criação do Conhecimento Organizacional	9
2.2 Gestão do Conhecimento	12
2.2.1 Gestão do Conhecimento Centrada em Produtos	14
2.2.2 Gestão do Conhecimento Centrada em Processos.....	15
2.2.3 A Gestão do Conhecimento e os Processos de Negócio	19
2.3 Gestão do Conhecimento Orientada a Processos de Negócio	20
2.3.1 BPO-KM	21
2.3.2 PROMOTE.....	21
2.3.3 CommonKADS	22
2.4 Processos Intensivos em Conhecimento.....	23
2.4.1 Formas de Representação.....	27
2.4.2 Elementos.....	29
2.5 Conclusão	31

Capítulo 3 – O Método <i>Story Mining</i>.....	33
3.1 Introdução e Trabalhos Relacionados	33
3.2 Mineração de Textos.....	35
3.2.1 Conceitos Básicos	37
3.3 O Método <i>Story Mining</i>	43
3.3.1 Fases da Mineração de Histórias.....	45
3.3.2 Solução Proposta pelo <i>Story Mining</i>	47
Capítulo 4 – Proposta de Solução	52
4.1 Requisitos da Proposta	52
4.2 Fases do Método.....	53
4.2.1 Selecionar Amostra de E-mails	54
4.2.2 Extrair e Estruturar Dados Relevantes	55
4.2.3 Minerar Conteúdo Textual dos E-mails	57
4.2.4 Minerar Dados Relevantes.....	61
4.2.5 Representar Processo Intensivo em Conhecimento.....	64
Capítulo 5 – Estudo de Caso e Análise de Resultados	71
5.1 Metodologia.....	71
5.2 Estudo Exploratório – Processo “Importar Mídias”.....	73
5.2.1 Descrição do Cenário	73
5.2.2 Execução.....	74
5.2.3 Análise dos Resultados.....	78
5.3 Estudo de Caso – Processo “Monitorar Envio de Notícias”	83
5.3.1 Descrição.....	83
5.3.2 Execução.....	85
5.3.3 Análise dos Resultados.....	86
Capítulo 6 – Conclusões e Trabalhos Futuros	94
6.1 Contribuições Realizadas.....	99
6.2 Limitações da Pesquisa.....	99
6.3 Trabalhos Futuros	100
Referências Bibliográficas	101

Índice de Figuras

Figura 2.1 – Espiral do Conhecimento (NONAKA, 1994)	11
Figura 2.2 – Facetas da Gestão do Conhecimento (Adaptada de ABECKER, 2004).....	13
Figura 2.3 - Apoio de Software à Gestão do Conhecimento Centrada em Produto e Processo	18
Figura 3.1 – Visão Macro do Processo de Mineração de Textos (GONÇALVES, 2010)	40
Figura 3.2 – Fases do Método <i>Story Mining</i> (GONÇALVES, 2010)	44
Figura 3.3 – Texto Gerado a partir da Análise Sintática (GONÇALVES, 2010)	46
Figura 3.4 – Exemplo de Resultado Gerado pela Fase de Mineração de Histórias.....	47
(GONÇALVES, 2010).....	47
Figura 3.5 – Workflow de Mineração de Texto do <i>Story Mining</i> (GONÇALVES, 2010)	48
Figura 4.1 – Fases do Método para Descoberta de Processos Intensivos em Conhecimento..	53
Figura 4.2 – Estrutura do XSD após Extração	56
Figura 4.3 – Workflow Científico Adaptado do Método <i>Story Mining</i>	59
Figura 4.4 – Estrutura do XSD após Mineração de Texto (<i>Story Mining</i>).....	61
Figura 4.5 – Visão Básica Expandida da Estrutura do Mapa Mental.....	68
Figura 4.6 – Estrutura do XSD Utilizado pelo <i>FreeMind</i>	69
Figura 4.7 – Processo Intensivo em Conhecimento em um Navegador Web	70
Figura 5.1 – Exemplo de Arquivo XML Gerado pela Fase de Extração	75
Figura 5.2 – Exemplo de Arquivo XML Gerado pela Fase de Mineração de Texto	75
Figura 5.3 – Exemplo de Arquivo XML Gerado pela Fase de Representação.....	76
Figura 5.4 – Representação do Processo Importar Mídias	77
Figura 5.5 – Script da Entrevista de Estudo de Caso	87

Índice de Tabelas

Tabela 2.1 - Principais Características entre as Abordagens de Produto e Processo (MENTZAS et al., 2002).....	17
Tabela 2.2 – Atributos de um Processo Intensivo em Conhecimento (EPPLER et al., 1999) ...	24
Tabela 5.1 – Relacionamento entre Papéis no Processo Monitorar Envio de Notícias	90

Capítulo 1 – Introdução

Este capítulo tem por objetivo apresentar um breve resumo desta dissertação, incluindo a sua relevância, justificativa, objetivos, limitações, procedimento metodológico e organização estrutural.

1.1 Relevância

Segundo DRUCKER (1993), com o crescimento da competição, o conhecimento apresenta-se como uma variável crítica para o sucesso. Intensas transformações sociais, econômicas e tecnológicas vêm redesenhando a atividade produtiva, ao colocar o conhecimento como principal recurso na criação de vantagens competitivas. Para preservar, compartilhar e reutilizar este patrimônio, é necessário que sejam adotadas estratégias apropriadas para apoiar a Gestão do Conhecimento.

STEELS (1993) define que os objetivos da Gestão do Conhecimento em uma organização devem consistir em promover o crescimento, comunicação e preservação do conhecimento organizacional, o que engloba gerenciar os recursos do conhecimento para facilitar o seu acesso e reuso.

Visando atender a estes objetivos, ABECKER (2004) introduz uma abordagem centrada em processos à Gestão do Conhecimento, sendo esta definida como uma forma de promover, motivar, nortear e guiar o processo de conhecer, descartando as tentativas de capturar e distribuir conhecimento. Esta visão permite compreender que a Gestão do Conhecimento trata de um processo de comunicação social, o qual deve ser apoiado por ferramentas colaborativas.

1.2 Problema

Existem, na literatura, algumas iniciativas voltadas à área de pesquisa da Gestão do Conhecimento Orientada a Processos, cujo principal objetivo consiste na representação de novos conhecimentos a partir da modelagem e execução de Processos de Negócio organizacionais, como em *BPO-KM* (HEISIG, 2003), *PROMOTE* (HINKELMANN, 2003), *CommonKADS* (SCHREIBER, 2000), entre outros.

GRONAU (2004) afirma que o principal objetivo, em longo prazo, da Gestão do Conhecimento Orientada a Processos consiste em identificar, modelar, analisar e aperfeiçoar os Processos Intensivos em Conhecimento. Contudo, as propostas existentes ainda não apresentam meios para a descoberta automática de Processos Intensivos em Conhecimento. HOFFMANN (2002) afirma que o apoio da Tecnologia da Informação e Comunicação (TIC) aos Processos Intensivos em Conhecimento não é ainda adequado.

Por esta razão, o problema identificado por esta pesquisa resume-se a: como descobrir automaticamente Processos Intensivos em Conhecimento?

1.3 Hipótese

ABECKER (2004) defende que, no ponto de vista da Gestão do Conhecimento Centrada a Processos, o conhecimento é mantido muito próximo aos indivíduos que o utilizam e compartilhado com base nos contatos entre as pessoas. O principal propósito da TIC, neste caso, consiste em apoiar a comunicação do conhecimento, e não apenas o seu armazenamento, coordenando e apoiando a colaboração.

Exemplos de ferramentas e sistemas de apoio à Gestão do Conhecimento Centrada em Processos envolvem todos os tipos de CSCW (*Computer Supported Collaborative Work* ou Trabalho Colaborativo Apoiado por Computador), entre elas as tecnologias de comunicação síncrona e assíncronas, como: chats, e-mails, videoconferências, grupos de discussão, etc.

HOFFMAN (2002) afirma que os Processos Intensivos em Conhecimento envolvem fortemente a socialização e a troca informal de conhecimento. Por essa razão, ferramentas colaborativas podem ser consideradas importantes fontes de informação, devido à flexibilidade inerente oferecida durante a execução destes processos.

Portanto, a hipótese desta pesquisa defende que: se for criado um método que permita a mineração de textos não estruturados a partir de ferramentas colaborativas, como e-mails, então será possível descobrir automaticamente Processos Intensivos em Conhecimento.

No escopo de atuação desta pesquisa será adotado o e-mail como ferramenta colaborativa a partir da qual serão extraídas as amostras de dados, visto que e-mails integram a cultura diária de todas as organizações.

1.4 Justificativa

ABECKER (2004) afirma que o ponto central dos Processos Intensivos em Conhecimento são os passos de decisão que requerem julgamentos pessoais baseados em conhecimento, criatividade e experiências dos participantes envolvidos.

VON HAGEN *et.al.* (2005) definem os Processos de Negócio Intensivos em Conhecimento como processos caracterizados pela forte dependência do conhecimento embutido nas pessoas que os executam e nas tarefas e atividades que os compõem.

EPPLER *et al.* (1999) defendem que o conhecimento embutido nos Processos Intensivos em Conhecimento é aplicado durante a execução de suas atividades. Logo, a descoberta deste conhecimento permite a possível influência em futuras decisões.

Dessa forma, acredita-se que a intensidade de conhecimento embutida nestes processos constitui rica fonte de conhecimento organizacional, o que torna a descoberta de informações relevantes aos Processos Intensivos em Conhecimento, um apoio fundamental à Gestão do Conhecimento organizacional.

1.5 Objetivos

O objetivo geral dessa dissertação é propor um método que torne possível a descoberta automática de informações relevantes aos Processos Intensivos em Conhecimento, que são fontes ricas de informações cruciais para as organizações, a partir da troca informal de conhecimento existente em ferramentas colaborativas, como e-mails.

Para alcançar este objetivo, os seguintes objetivos específicos foram definidos:

- Criar uma solução que apoie a extração de conhecimento a partir da mineração de textos provenientes de ferramentas colaborativas.
- Adaptar um framework de mineração de textos não estruturados existente, expandindo sua aplicabilidade.
- Aplicar o método proposto em um ambiente organizacional com cenário real.

1.6 Delimitações de Abrangência

No que tange às limitações desta pesquisa, esta irá apresentar, para fins de fundamentação teórica, os conceitos fundamentais relacionados a Processos de Negócio, Gestão do Conhecimento, Gestão do Conhecimento Orientada a Processos e Processos Intensivos em Conhecimento. Tais conceitos têm por função estabelecer o embasamento teórico necessário ao desenvolvimento do método.

Como o principal foco deste trabalho trata da elaboração de um método para descoberta de informações relevantes aos Processos Intensivos em Conhecimento através da mineração de textos não estruturados provenientes de ferramentas colaborativas, serão conceituados os métodos e técnicas aplicados para este objetivo. Neste estudo serão abordados conceitos de mineração de texto necessários ao objetivo de possibilitar o entendimento do trabalho realizado, não estando, portanto no escopo do trabalho métodos, técnicas e ferramentas específicas à mineração de textos em geral.

Como premissa de aplicação do método proposto, foi importante a seleção de amostras de dados satisfatórias, levando em consideração a quantidade e qualidade dos e-mails existentes, de forma que fosse possível a descoberta de informações significativas, evitando ruídos e melhorando a qualidade do resultado apresentado.

1.7 Procedimento Metodológico

A pesquisa realizada pode ser classificada quanto à sua natureza como **Pesquisa Aplicada**, pois objetiva gerar conhecimento para aplicação prática dirigida à solução do problema em questão.

Quanto aos seus objetivos como Pesquisa Exploratória, pois visa proporcionar maior familiaridade com o problema apresentado, uma vez que ainda há pouca discussão em torno deste tema.

Quanto à sua abordagem como Pesquisa Qualitativa, pois há um vínculo indissociável entre o mundo objetivo e a subjetividade do sujeito que não pode ser traduzido em números. Nesta pesquisa, o cenário real de aplicação do método é a fonte direta para coleta de dados, sendo estes descritivos e analisados indutivamente.

Ainda quanto a procedimento técnico como Estudo de Caso, uma vez que será necessária a aplicação do método proposto em uma empresa de Mídia e Entretenimento, envolvendo o estudo profundo e exaustivo dos resultados de maneira que se permita o seu amplo e detalhado conhecimento.

Quanto à coleta de dados como Entrevista Estruturada, onde serão obtidas informações de todos os participantes sobre o resultado gerado pelo método proposto. Tais informações serão levantadas com base em um roteiro previamente estabelecido.

Quanto ao método científico será adotado um Método Indutivo de forma a provar que se todas as premissas apresentadas forem verdadeiras, a conclusão é provavelmente verdadeira.

1.8 Organização Estrutural

Este trabalho foi dividido em seis capítulos, sendo este primeiro a Introdução.

O **Capítulo 2** introduz as fundamentações teóricas necessárias para estabelecer um embasamento conceitual que permita o desenvolvimento do método proposto, sendo estes conceitos de Processos de Negócio, Gestão do Conhecimento, Gestão do Conhecimento Orientada a Processos e Processos Intensivos em Conhecimento.

O **Capítulo 3** discute o método *Story Mining*, uma abordagem que associa o uso de narrativas colaborativas (*Group Storytelling*) com técnicas de mineração de texto para apoio à elicitación de Processos de Negócio, apresentando métodos e técnicas de mineração de textos não estruturados necessários para alcançar tais objetivos.

O **Capítulo 4** apresenta a proposta de solução para o método de descoberta de Processos Intensivos em Conhecimento através da mineração de textos não estruturados provenientes de ferramentas colaborativas (e-mails), contribuição principal desta pesquisa.

O **Capítulo 5** apresenta a avaliação e instanciação do método proposto com base no cenário definido para o estudo de caso, um Processo Intensivo em Conhecimento executado em uma empresa de Mídia e Entretenimento.

O **Capítulo 6** estabelece um resumo de toda a dissertação e realiza uma análise das contribuições, conclusões, assim como apresenta possíveis trabalhos futuros de pesquisa.

Capítulo 2 – Fundamentação Teórica

Este capítulo tem como objetivo apresentar os conceitos necessários para a compreensão adequada das ideias elaboradas nesta pesquisa, provendo o embasamento teórico necessário.

2.1 Conhecimento

NONAKA (1994) apresenta o conhecimento como o fator responsável por aumentar a capacidade de um indivíduo para tomar uma ação, de acordo com suas habilidades, competências e atividades intelectuais. Também define Conhecimento como toda a informação retida na mente de um indivíduo, personalizada ou relacionada a fatos, conceitos, interpretações, ideias, observações e julgamentos.

DAVENPORT & PRUSAK (2000) definem o conhecimento como um conjunto de experiências, valores e informações contextuais e *insights* de especialistas que fornece uma forma para avaliar e incorporar experiências e informação. O Conhecimento é originado a partir da mente dos indivíduos e, 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.

SCHREIBER *et al.* (2000) resume o Conhecimento em um conjunto de informações usadas por organizações ou indivíduos para a execução de ações visando à criação de novas informações.

Além destas definições, pode ser encontrado na literatura um volume imenso de artigos e teses que propõem a discussão exaustiva do Conhecimento aplicado às mais diversas perspectivas, como: Cognição, Psicologia Cognitiva, Ciências Sociais e Pedagógicas, Teoria Organizacional, Aprendizagem Organizacional, Psicologia Organizacional, Inteligência Organizacional, Memória Organizacional, etc.

Em todas as abordagens são discutidas diferentes questões do conhecimento, como: Conhecimento Tácito *versus* Conhecimento Explícito, Conhecimento Coletivo, Representação do Conhecimento, Tipos de Conhecimento Organizacional, entre outros. Por esta razão, para uma contextualização mais objetiva do Conhecimento, serão apresentadas as suas principais características sob um prisma especificamente voltado para sua aplicação em Sistemas de Informação.

2.1.1 Conhecimento e os Sistemas de Informação

SCHEIR (2002) apresenta uma relação contendo as principais características do Conhecimento quando aplicado ao contexto de Sistemas de Informação, sendo estas:

- **Orientado a Ação:** conhecimento é criado ativamente durante a interação de um indivíduo com o seu ambiente;
- **Subjetividade:** conhecimento é criado, adquirido e ativado no contexto de condições ambientais específicas;
- **Dependência Social:** conhecimento é criado a partir de contatos sociais e relacionamentos entre indivíduos;
- **Grau de Consciência:** conhecimento nem sempre é consciente quando aplicado para a resolução de problemas.

Com base nesta lista de características é possível compreender que o conhecimento, dado a sua subjetividade, não é facilmente tratado por sistemas computacionais. Para que isso seja possível, é fundamental que as soluções inovadoras reflitam estas características através de suas abordagens e arquiteturas.

SCHEIR (2002) resume os pontos essenciais do conhecimento que necessitam estar refletidos em Sistemas de Informação:

- Conhecimento é orientado a propósito e orientado à resolução de problemas;
- Conhecimento consiste de informação contextualizada e interligada;
- Conhecimento está limitado nos modelos internos dos indivíduos.

As características do conhecimento apresentadas permitem a visão dos principais pontos a serem considerados em pesquisas na área de Sistemas de Informação.

2.1.2 Criação do Conhecimento Organizacional

NONAKA E TAKEUCHI (1995) classificam o Conhecimento em dois tipos distintos: Conhecimento Explícito e Conhecimento Tácito. Pode ser considerado Explícito, todo o conhecimento formalizado em diversos meios, como Documentos, Livros, Procedimentos, Repositório de Dados, entre outros, de forma que possa ser transmitido através de linguagem formal e sistemática. Por outro lado, é considerado Tácito, todo o conhecimento adquirido durante a experiência de vida, da educação e do aprendizado, estando este presente na mente do indivíduo.

Apesar desta distinção, o conhecimento tácito e o conhecimento explícito não são entidades totalmente separadas, e sim mutuamente complementares, uma vez que interagem um com o outro e realizam trocas nas atividades criativas dos seres humanos. Por essa razão, cabe às organizações apoiar os indivíduos criativos e propiciar contextos para que novos conhecimentos sejam criados (NONAKA & TAKEUCHI, 1995).

A criação do conhecimento organizacional deve ser compreendida como um processo que amplifica organizacionalmente o conhecimento criado pelos indivíduos e o consolida como parte da sua rede de conhecimentos.

A criação de conhecimento organizacional resulta da conversão de conhecimento tácito em conhecimento explícito. As conversões do conhecimento entre esses dois formatos constituem a essência da abordagem teórica da criação do conhecimento criada por NONAKA & TAKEUCHI (1995). O modelo proposto parte do princípio que os conhecimentos tácito e explícito fazem parte da organização e, portanto, ocorre a conversão do conhecimento quando estes interagem entre si. Este processo pode ocorrer de quatro formas distintas:

- **Socialização:** é a transferência de conhecimento tácito para tácito, por meio do compartilhamento de experiências e da observação, através da troca de e-mails, conversas, aulas, reuniões, palestras, etc.;
- **Externalização:** processo de conversão de conhecimento tácito em explícito através do uso de palavras, imagens, metáforas, analogias e hipóteses, além da dedução, indução e abdução;
- **Combinação:** é a análise conjunta de distintos conhecimentos explícitos, como documentos, memorandos e banco de dados, de forma a gerar, a partir destes, um novo conhecimento explícito;
- **Internalização:** é a conversão de conhecimento explícito em conhecimento tácito.

A proposta de NONAKA (1994) é que um modelo organizacional de conhecimento passe pela interação dinâmica entre os diferentes modos de conversão de conhecimento, de forma que a criação do conhecimento ocorra baseada na construção de conhecimentos tácito e explícito e na operação entre os dois através da internalização e externalização. Ou seja, para a criação do conhecimento organizacional, faz-se necessária a aplicação das quatro formas de conversão de conhecimento através de um ciclo contínuo, conforme descrito pela conhecida Espiral do Conhecimento, ilustrada na Figura 2.1:

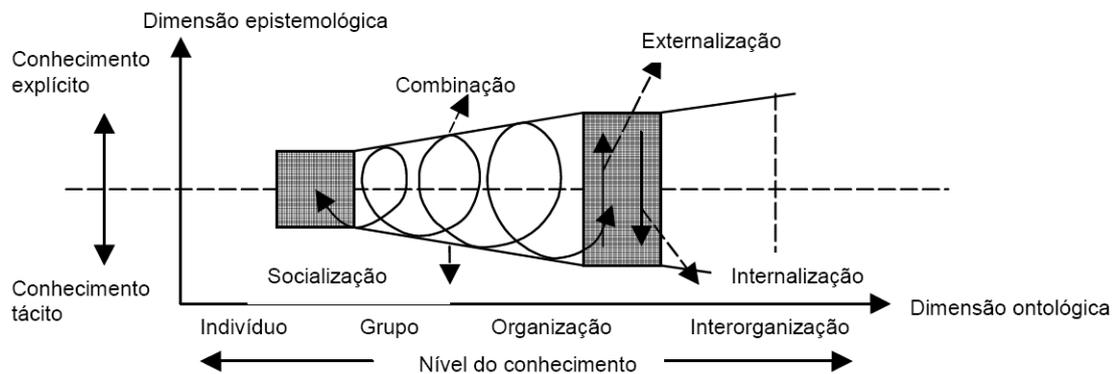


Figura 2.1 – Espiral do Conhecimento (NONAKA, 1994)

Com base na figura acima, é possível afirmar que a criação do conhecimento é sempre iniciada a partir da socialização, uma vez que as interações sociais facilitam o compartilhamento de experiências e perspectivas dos membros das equipes. A seguir, inicia-se a externalização, através de sucessivas rodadas de "diálogos". Com base nestes diálogos, entidades articulam seu conhecimento tácito para que os demais indivíduos possam entender. O conhecimento criado através da externalização pode ser combinado com o conhecimento existente para, posteriormente, ser refinado e estender a base de conhecimento ou memória organizacional.

Este processo ocorre iterativamente, possibilitando o crescimento e evolução do conhecimento de forma concreta. Através desta experimentação de aprendizado, é que a internalização, verdadeiramente, acontece. As interações entre conhecimento tácito e explícito tendem a se tornar maiores e mais rápidas à medida que mais atores da organização são envolvidos. Assim, a criação do conhecimento organizacional pode ser vista como uma espiral ascendente, que inicia no indivíduo e move-se cada vez mais ao nível da organização.

2.2 Gestão do Conhecimento

Com o crescimento da competição, o conhecimento apresenta-se como uma variável crítica para o sucesso. A necessidade de novas abordagens e respostas aos paradigmas impostos às organizações na atual “era do conhecimento” é patente. Intensas transformações sociais, econômicas e tecnológicas vêm redesenhando a atividade produtiva, ao colocar o conhecimento como principal recurso na criação de vantagens competitivas. As organizações dependem da acumulação e aplicação do conhecimento para criar valor econômico (DRUCKER, 1993).

Segundo DAVENPORT & PRUSAK (2000), o conhecimento é uma arma para o sucesso competitivo. Portanto, para se manterem competitivas, as organizações precisam gerenciar o conhecimento, assim como seus ativos intangíveis. Segundo BUSCH (2008), a globalização força a inovação e, conseqüentemente, esta aponta para a necessidade de Gestão do Conhecimento.

NONAKA E TAKEUCHI (1995) afirmam que, no contexto de desenvolvimento de novos produtos, as empresas bem sucedidas se caracterizam por sua habilidade de consistentemente criar novos conhecimentos disseminando-os e incorporando-os rapidamente em novos produtos. Portanto, faz-se mister a elaboração de novas estratégias e métodos que auxiliem a criação, aquisição, gerenciamento e, principalmente, materialização do conhecimento na forma de inovação em produtos e serviços.

Para atender a estes objetivos foi criada a área de Gestão do Conhecimento. Nesta seção serão resumidos alguns conceitos fundamentais que permitam a sua compreensão.

Um relatório tecnológico elaborado pela OVUM (1998) caracteriza a Gestão do Conhecimento como uma tarefa responsável pelo desenvolvimento e exploração dos recursos tangíveis e intangíveis de conhecimento.

ABECKER (2004) define que um dos principais objetivos da Gestão do Conhecimento consiste em adquirir, processar e criar conhecimento organizacional acessível, de forma sistemática, visando prover as melhores decisões e estar mais bem preparado para o futuro, através da identificação, desenvolvimento e provisão do conhecimento relevante ao sucesso da organização.

DAVENPORT & PRUSAK (2000) descrevem a Gestão do Conhecimento como uma iniciativa formal e estruturada para prover a criação, distribuição e uso do conhecimento em uma organização, através de um processo formal capaz de transformar conhecimento corporativo em valor organizacional.

MAURER (1999) defende que, sob a perspectiva da Inteligência Artificial, o objetivo principal da Gestão do Conhecimento deve ser nortear e aumentar o conhecimento dos indivíduos e garantir que este seja facilmente compartilhado com todos e permaneça ativo, mesmo que as pessoas envolvidas não estejam disponíveis.

STEELES (1993) define que os objetivos da Gestão do Conhecimento em uma organização devem consistir em promover o crescimento, comunicação e preservação do conhecimento organizacional, o que engloba gerenciar os recursos do conhecimento para facilitar o seu acesso e reuso.

Além destas definições, poderiam ser citadas ainda diversas outras relacionadas à Gestão do Conhecimento. No entanto, acredita-se que estas definições oferecem aspectos relevantes para a compreensão desta pesquisa. A figura 2.2 resume as principais facetas referentes à área de Gestão do Conhecimento:

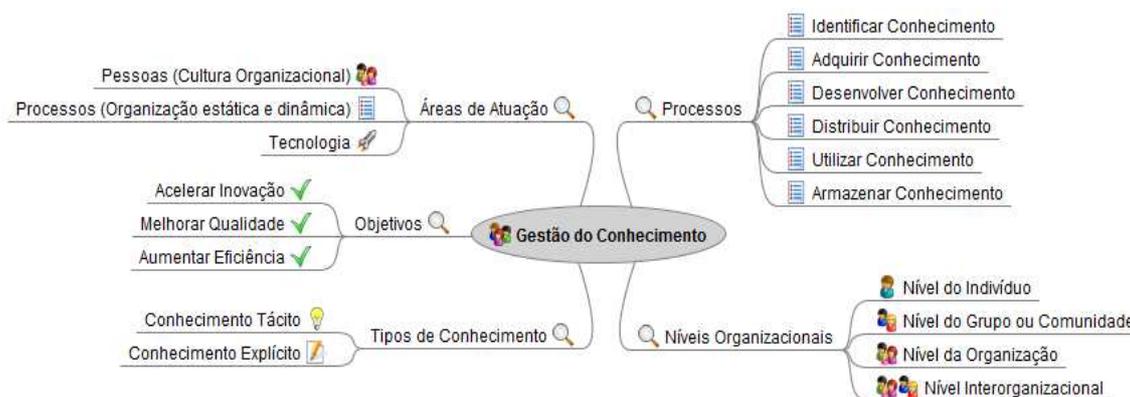


Figura 2.2 – Facetas da Gestão do Conhecimento (Adaptada de ABECKER, 2004)

ABECKER (2004) resume a Gestão do Conhecimento como sendo:

- uma abordagem holística e estruturada;
- utilizada para prover a melhoria sustentável no manuseio do conhecimento tácito e explícito das organizações;
- aplicada em todos os níveis (individual, grupal, organizacional e interorganizacional);
- ideal para alcançar um ou mais objetivos estratégicos organizacionais, como diminuir custos, melhorar a qualidade, produzir inovações, aumentar a satisfação dos clientes, entre outros.

Ao analisar as diversas aplicações relacionadas à Gestão do Conhecimento, KÜHN E ABECKER (1997) classificam as abordagens existentes em dois tipos distintos: Centradas em Processos e Centradas em Produtos.

2.2.1 Gestão do Conhecimento Centrada em Produtos

A abordagem Centrada em Produtos visa a captura, distribuição e gestão do conhecimento propriamente dito, mantendo o seu foco em produtos e artefatos que contenham ou representem conhecimento. Esta abordagem visa basicamente a gestão de documentos, ou seja, a criação, armazenamento e reuso de documentos através da utilização de Sistemas de Informação de Memória Organizacional (ABECKER *et al.*, 1998).

Segundo WEBER *et al.* (2001), adotar o conhecimento como produto visa tratar o conhecimento como uma entidade, de forma que possa ser separada das pessoas que o criam e utilizam. O seu objetivo específico consiste em identificar documentos que contenham conhecimento explícito, como relatórios, apresentações, artigos, entre outros, e armazená-los em repositórios, a partir dos quais, estes documentos possam ser facilmente recuperados. Exemplos de ferramentas e sistemas voltados para esta abordagem incluem: repositórios de melhores práticas e lições aprendidas, taxonomias de conhecimento e estruturas formais de conhecimento para portais semânticos.

Tipos comuns de projetos que representam esta abordagem são:

- **Repositórios de Conhecimento:** os também chamados Repositórios de Conhecimento Externo tratam da busca automática de informação, a partir de *Clipping Services* (canais de busca de informação), capazes de rotear desde artigos a executivos, de competitividade avançada a sistemas de inteligência de clientes, através de sistemas baseados em técnicas de Extração de Informação (*Information Extraction*) que visam detectar eventos a partir de grandes corpora de texto.
- **Conhecimento Interno Estruturado:** trata do armazenamento de documentos estruturados, como relatórios de pesquisa, técnicas corporativas e métodos.
- **Conhecimento Interno Informal:** trata do armazenamento de conhecimento a partir de Banco de Dados de discussões, como Repositórios de e-mails, Fóruns e Repositórios de Lições Aprendidas.

2.2.2 Gestão do Conhecimento Centrada em Processos

A abordagem centrada em processos pode ser definida como uma forma de promover, motivar, nortear e guiar o processo de conhecer, descartando as tentativas de capturar e distribuir conhecimento. Esta visão permite compreender que a Gestão do Conhecimento trata de um processo de comunicação social, o qual deve ser apoiado por ferramentas colaborativas (ABECKER, 2004).

Nesta abordagem, o conhecimento é mantido muito próximo à pessoa que o utiliza e compartilhado com base nos contatos entre as pessoas. O principal propósito da Tecnologia da Informação e Comunicação, neste caso, consiste em apoiar a comunicação do conhecimento, e não apenas o seu armazenamento, coordenando e apoiando a colaboração.

Exemplos de ferramentas e sistemas de apoio à Gestão do Conhecimento Centrada em Processos, incluindo todos os tipos de CSCW (*Computer Supported Collaborative Work* ou Trabalho Colaborativo Apoiado por Computador), são:

- **Tecnologias de comunicação síncrona e assíncronas**, como: chats, videoconferências, grupos de discussão, lista de e-mails, etc.
- **Sistemas de apoio à coordenação do trabalho colaborativo**, como: sistemas de gestão de processos de negócio, ferramentas de gestão de projetos Web, *workspaces* eletrônicas compartilhadas, entre outros.
- **Sistemas para a colaboração otimizada**, como: tecnologia para criação de documentos hipertexto distribuídos, sistemas de apoio à Decisão em Grupo, ferramentas de reunião e busca de informações colaborativas.

Considerar o conhecimento como um processo permite o surgimento e desenvolvimento de novos conceitos, como as Comunidades, que se tornam peças-chave para a distribuição e, conseqüente, nivelamento do conhecimento. Por esse motivo, algumas empresas que adotam a abordagem da Gestão do Conhecimento Centrada em Processos, mantêm seu foco na criação das chamadas Comunidades de Prática ou Comunidades de Interesse. Estas comunidades consistem em grupos auto-organizáveis, onde a comunicação flui naturalmente entre os envolvidos, uma vez que compartilham de interesses e práticas em comum. Através da criação destas comunidades, é possível gerar e compartilhar conhecimento (WENGER, 1998).

Nesta abordagem, a ênfase está em prover acesso e transferência do conhecimento entre os indivíduos. Por essa razão, essas iniciativas dependem fortemente da qualidade da participação de gerenciamento organizacional, em prol de incentivar a confiança entre os membros dos grupos, facilitar a troca de experiências, cultivar um clima positivo e prover papéis apropriados que facilitem o trabalho nestas comunidades (WENGER, 1998).

A dicotomia existente entre as abordagens centradas em produto e processo tornam-se evidentes em diversas iniciativas da Gestão do Conhecimento e, conseqüentemente, têm sido descobertas, redescobertas e analisadas sob diversas perspectivas na literatura científica. Na Tabela 2.1, MENTZAS *et al.* (2002) resumem as principais diferenças entre as abordagens centradas em produto e processo para a Gestão do Conhecimento:

	Conhecimento como Produto	Conhecimento como Processo
Visão	Conhecimento pode ser representado como algo que pode ser localizado e manipulado como um objeto independente. Ênfase na captura, distribuição e métricas de conhecimento.	Só é possível promover, motivar, encorajar, nortear e guiar o processo de conhecer. Nesta visão, a ideia de capturar e distribuir conhecimento torna-se sem sentido.
Foco	Produtos e artefatos que contenham ou representem conhecimento. Normalmente, implica em gerenciar documentos e dados, promovendo criação, armazenamento e reuso através de repositórios computacionais.	Gestão do Conhecimento como um processo de comunicação social, a qual pode ser sempre aperfeiçoada através de ferramentas de colaboração e cooperação.
Estratégia	Explorar conhecimento organizado, padronizado e reutilizável.	Incentivar o canal de comunicação entre indivíduos e equipes para a troca de experiências.
Foco da GC	Integrar indivíduos ao conhecimento reutilizável e codificado.	Facilitar o diálogo para a troca de conhecimento.
Foco do RH	Realizar treinos em grupos. Recompensar pela utilização e contribuição em repositórios de conhecimento.	Incentivar treinos a partir da aprendizagem. Recompensar o compartilhamento de conhecimento com os demais.
Foco da TIC	Forte ênfase em Sistemas de Gestão de Documentos.	Ênfase moderada em Sistemas de Gerenciamento de Redes.
Tecnologias Adotadas	Repositórios de Documentos, busca avançada de informações, bancos de dados e mapas de conhecimento.	Grupos de discussão, e-mails, videoconferências e serviços de mensagens instantâneas.

Tabela 2.1 - Principais Características entre as Abordagens de Produto e Processo (MENTZAS *et al.*, 2002)

Obviamente, os desafios holísticos e efetivos da Gestão do Conhecimento devem considerar igualmente ambas as abordagens, de forma otimizada e sinérgica. Conforme apresentado anteriormente, sobre as práticas, ferramentas e casos de sucesso da Gestão do Conhecimento, é possível denotar que esta só pode ser aplicada a partir da utilização de tecnologias convencionais. A Figura 2.3 ilustra o apoio de software à Gestão do Conhecimento Centrada em Produto e Processo:

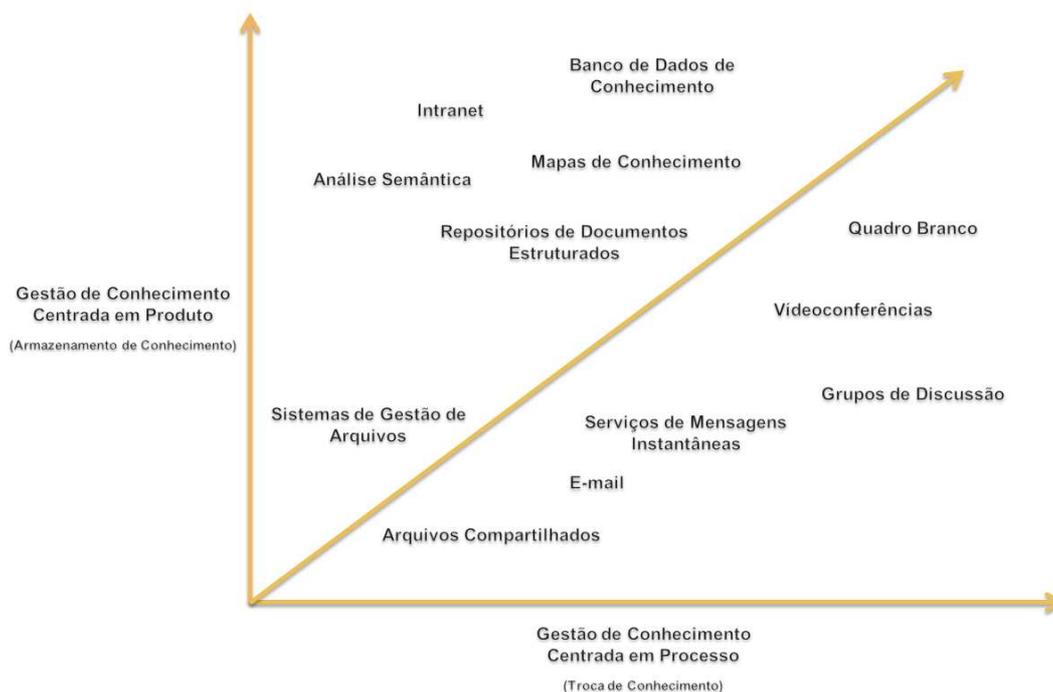


Figura 2.3 - Apoio de Software à Gestão do Conhecimento Centrada em Produto e Processo

Segundo ABECKER (2004), o advento de tecnologias ligadas à Intranet e Internet foi um dos principais apoiadores ao início da aplicação da Gestão do Conhecimento, já que permitiu a utilização de novas formas de comunicação eletrônica e colaboração à distância. O desenvolvimento de tecnologias poderosas para Busca de Informações, Análise e Classificação de Textos, foi outro facilitador, uma vez que tornou possível o manuseio altamente efetivo de conhecimento explícito, como fontes da Internet, arquivos corporativos e banco de dados de conhecimento, por exemplo, repositórios de Lições Aprendidas, concluindo, portanto, o papel fundamental da TI no apoio à Gestão do Conhecimento.

2.2.3 A Gestão do Conhecimento e os Processos de Negócio

EPPLER *et al.* (1999) introduzem o conceito de Agentes à definição de conhecimento, considerando estes, indivíduos, comunidades ou algoritmos de processamento de informação. Estes agentes processam informações relacionadas a atividades de processos de negócio, utilizando-a, efetivamente, em futuras decisões. Sendo assim, os agentes são capazes de armazenar e produzir informações, tal como, tomar decisões conforme o seu conhecimento. Estes agentes interagem entre si, segundo protocolos definidos, e são assinalados a um ou mais papéis do processo de negócio, tendo este, portanto, distintas responsabilidades e tarefas. Sob esta perspectiva, o conhecimento de um processo de negócio consiste na informação contextual presente nos agentes envolvidos, servindo como possível influência para decisões futuras. A partir desta definição de conhecimento em processos de negócio, EPPLER *et al.* (1999) apresentam as seguintes conclusões:

- O contexto do conhecimento consiste no propósito do processo, por exemplo, a elaboração de um novo produto em um Processo de Desenvolvimento de Produtos;
- A informação embutida no processo de negócio é produzida durante a execução de suas atividades, o que torna necessária a sua utilização consistente no processo;
- A possível influência para futuras decisões ocorre caso a informação contextual seja reproduzível pelo agente;

Em resumo, esta perspectiva defende a existência de conhecimento a partir das atividades executadas por Agentes em Processos de Negócio. De forma análoga à abordagem Centrada em Processos, a Gestão do Conhecimento apresenta-se como possível forma de apoio à descoberta de conhecimento através dos Processos de Negócio, provendo os métodos e ferramentas necessários. Portanto, para atender a este objetivo, foi criada uma subárea de pesquisa denominada Gestão do Conhecimento Orientada a Processos de Negócio.

2.3 Gestão do Conhecimento Orientada a Processos de Negócio

A Gestão do Conhecimento Orientada a Processos de Negócio (*Business Process Oriented Knowledge Management* ou BPOKM) consiste em uma subárea de pesquisa da Gestão do Conhecimento cujo foco é voltado para a descoberta do conhecimento organizacional existente em Processos de Negócio.

ABECKER (2002) fundamenta esta subárea nos modelos de mudanças dinâmicas existentes dentro das organizações, tendo como base, não apenas o conhecimento existente nos Processos de Negócio organizacionais, como também o conhecimento integrante da Gestão de Processos de Negócio e da Cultura Corporativa. Portanto, a Gestão do Conhecimento Orientada a Processos consiste em um campo de pesquisa que utiliza a modelagem de Processos de Negócio como base para a aplicação da Gestão do Conhecimento, podendo esta ocorrer em três cenários:

- Processos de Negócio como ponto de partida para a Gestão do Conhecimento;
- Gestão do Conhecimento com base na execução de processos;
- Processos de negócio como razão da Gestão do Conhecimento.

Para atender a esta área de pesquisa, existem algumas propostas na literatura que visam o apoio a estas práticas através da criação de métodos computacionais. ABECKER (2002) apresenta para estas propostas, a necessidade de se aplicar algumas ou todas as fases de projeto de sistemas convencionais, como: Planejamento, Análise, Implementação, Usabilidade e Evolução. Para classificar as distintas abordagens de pesquisa, ABECKER (2002) propõe a subdivisão destas fases em três camadas distintas:

- Gestão do Conhecimento Orientada a Processos Estratégicos;
- Projeto de Gestão do Conhecimento baseado em Processos de Negócio;
- Projeto de Gestão do Conhecimento baseado na Análise de Comunicação.

Na camada superior, a Gestão do Conhecimento Orientada a Processos Estratégicos consiste em uma perspectiva *top-down* responsável por derivar objetivos de conhecimento a partir de objetivos de negócio em longo prazo.

Na camada inferior, o Projeto de Gestão do Conhecimento baseado na Análise de Comunicação lida com aspectos da comunicação voltados ao trabalho de conhecimento, a fim de desenvolver métodos e ferramentas apropriadas.

Na camada intermediária, são propostos métodos e ferramentas para análise de processos de negócio adaptados, de forma a atender aos requisitos da Gestão do Conhecimento. Esta camada intermediária lida com métodos de modelagem derivados da Gestão de Processos de Negócio e Modelagem de Processos visando a identificação de potenciais melhorias.

Para atender a estes objetivos, existem algumas propostas na literatura, como: *BPO-KM*, *PROMOTE*, *CommonKADS*, entre outras.

2.3.1 BPO-KM

HEISIG (2003) define o método BPO-KM para análise e projeto de soluções orientadas a processo para a Gestão do Conhecimento. No primeiro passo desta abordagem, a auditoria da Gestão do Conhecimento analisa condições fundamentais, incluindo avaliação de sistemas de TI existentes, análise da Cultura Informacional e de Conhecimento e a determinação da demanda para a Informação e Conhecimento. O foco principal deste passo é a identificação do contexto de negócio. O passo seguinte analisa eventuais Processos Intensivos em Conhecimento existentes, para identificar qualidades, defeitos e possíveis melhorias. No último passo, são identificadas demandas relacionadas aos processos e suas respectivas tarefas.

2.3.2 PROMOTE

HINKELMANN (2003) apresenta outra abordagem existente, o método PROMOTE, que integra planejamento estratégico com a avaliação da Gestão do Conhecimento e Gestão de Processos de Negócio.

Esta abordagem cobre a análise, modelagem e execução de Processos Intensivos em Conhecimento. Para isso, estende métodos genéricos de Sistemas de Gestão de Processos de Negócio, incluindo decisões estratégicas, reengenharia e alocação de recursos e avaliação de desempenho. Os passos adicionais desta abordagem consistem em criar Conhecimento Corporativo, descobrir conhecimento a partir de processos, criar Processos de Conhecimento Operacional, apoiar a Memória Organizacional e avaliar Conhecimento Corporativo (HINKELMANN, 2003 *apud* GRONAU, 2004).

2.3.3 CommonKADS

SCHREIBER (2000) adapta a abordagem de Engenharia de Conhecimento, *CommonKADS*, à captura de Processos Intensivos em Conhecimento. Apesar de o seu foco ser um pouco distinto do apresentado por esta área, o seu processo de aquisição de conhecimento pode ser considerado aplicável, uma vez que as suas técnicas de aquisição de conhecimento incluem a elicitación, coleção, análise, modelagem e validação de conhecimento para a Engenharia de Conhecimento e Projetos de Gestão do Conhecimento. As técnicas de aquisição de conhecimento propostas foram desenvolvidas de forma a auxiliar a elicitación de conhecimento por parte dos especialistas.

Estes e outros métodos propostos auxiliam a Gestão do Conhecimento Orientada a Processos, uma vez que proveem passos fundamentais à descoberta de novos conhecimentos, a partir da análise de processos de negócio organizacionais. Neste contexto, é possível denotar a relevância dos Processos Intensivos em Conhecimento, também conhecidos como Processos de Negócio Intensivos em Conhecimento. Segundo GRONAU (2004), identificar, modelar, analisar e otimizar Processos Intensivos em Conhecimento deve ser o principal objetivo, a longo prazo, da Gestão do Conhecimento Orientada a Processos. A Gestão do Conhecimento e Processos de Negócio estão integrados e devem ser analisados como um todo.

2.4 Processos Intensivos em Conhecimento

EPPLER *et al.* (1999) apresentam uma definição completa do Conhecimento, sob a perspectiva de Processos de Negócio, como sendo o conhecimento de, sobre e derivado de um processo, o qual deve ser compartilhado como uma informação dependente de contexto capaz de auxiliar os agentes do processo a tomarem efetivas decisões futuras. No entanto, a análise de Processos de Negócio sob uma perspectiva de Conhecimento só é possível caso haja potencial existência de conhecimento relacionado ao processo, ou seja, caso o conhecimento realmente faça a diferença para sua execução, como em processos de pesquisa, desenvolvimento de produtos, planejamento, entre outros.

EPPLER *et al.* (1999) propõe a classificação de Processos de Negócio baseada na avaliação da intensidade do conhecimento existente e da sua complexidade. Por essa razão, distingue-os em quatro classes de processo:

- Processos de baixa complexidade e fraca intensidade de conhecimento
(Classe 1);
- Processos de alta complexidade e fraca intensidade de conhecimento
(Classe 2);
- Processos de alta complexidade e forte intensidade de conhecimento
(Classe 3);
- Processos de baixa complexidade e forte intensidade de conhecimento
(Classe 4).

Dada as possíveis classificações de processo, conforme suas respectivas complexidade e intensidade de conhecimento, no contexto desta pesquisa, serão aprofundados os conceitos relacionados aos processos de Classes 3 e 4, dada a forte intensidade de conhecimento que deseja-se observar.

Para que seja possível avaliar a intensidade de conhecimento existente em um Processo de Negócio, ou seja, para que um processo possa ser analisado sob uma perspectiva de conhecimento, são necessários atributos que auxiliem esta classificação. *EPPLER et al.* (1999) descrevem seis atributos que necessitam ser considerados ao avaliar a intensidade de conhecimento de um Processo de Negócio. A Tabela 2.2 descreve estes atributos:

Atributos	Descrição	Intensidade de Conhecimento
Contingência	As atividades são contingentes devido a diversas eventualidades e são diretamente impactadas por questões de ambiente.	Forte
	As atividades são definidas e mapeadas e não são diretamente impactadas por questões de ambiente.	Fraca
Escopo de Decisão	O agente dispõe de diversas possibilidades de como deve proceder durante as decisões do processo.	Forte
	O agente não pode variar durante a execução de suas atividades.	Fraca
Inovação do Agente	O agente deve resolver problemas de forma criativa e inovadora.	Forte
	O processo não requer criatividade nem inovação.	Fraca
Tempo de Vida	O conhecimento do agente rapidamente torna-se obsoleto, uma vez que o conhecimento relacionado ao processo atualiza-se constantemente.	Forte
	O conhecimento do agente, uma vez estabelecido, é relevante por um longo período de tempo.	Fraca
Impacto do Agente	O agente possui forte influência no resultado do processo.	Forte
	O agente possui pouca influência no resultado do processo.	Fraca
Tempo de Aprendizagem	O agente necessita de um longo período de tempo para adquirir as habilidades necessárias para desempenhar suas tarefas.	Forte
	As habilidades necessárias para que os agentes executem suas atividades podem ser rapidamente adquiridas.	Fraca

Tabela 2.2 – Atributos de um Processo Intensivo em Conhecimento (*EPPLER et al.*, 1999)

HEISIG (2003) ainda avalia a intensidade de conhecimento como diretamente proporcional às variabilidades e exceções existentes nos processos. Consequentemente, conclui afirmando que um processo pode ser caracterizado como intensivo em conhecimento quando é impossível ou parcialmente possível aplicar, a este processo, métodos convencionais de Redesenho de Processos de Negócio. DAVENPORT *et al.* (1996) identificam Processos Intensivos em Conhecimento através da diversidade e incerteza de suas entradas e saídas.

Baseado em todas as características apresentadas, é possível afirmar que um processo é intensivo em conhecimento quando, para sua execução, é mandatório o preenchimento de todos os requisitos de conhecimento necessários por parte dos Participantes envolvidos. Além destas definições, existem na literatura diversas outras tentativas para definir um Processo Intensivo em Conhecimento.

EPPLER (1999) afirma que em Processos Intensivos em Conhecimento, o conhecimento contribui significativamente aos valores agregados. Nestes processos, inovação e criatividade são fundamentais, uma vez que os seus participantes possuem plena liberdade para decisões autônomas.

DAVENPORT *et al.* (1996) identificam que o fluxo de eventos em Processos Intensivos em Conhecimento não é claramente conhecido, uma vez que este pode evoluir durante a sua execução.

HEISIG (2003) defende que em Processos Intensivos em Conhecimento os Participantes possuem experiências distintas e aplicam conhecimento sob diversos domínios em diferentes níveis de especialidade.

EPPLER (1999) afirma que o tempo de vida do conhecimento envolvido em Processos Intensivos em Conhecimento é normalmente muito curto. Sendo assim, o conhecimento presente nestes processos torna-se obsoleto rapidamente, o que exige muito tempo disponível dos participantes para adquiri-lo.

DAVENPORT & PRUSAK (2000) caracteriza os Processos Intensivos em Conhecimento como sendo não estruturados, uma vez que não seguem regras de trabalho estruturadas e, dificilmente, apresentem métricas para avaliar o seu sucesso.

HAMEL (1990) defende que Processos Intensivos em Conhecimento devem ser o cerne dos processos organizacionais, de forma a produzir ou criar novos conhecimentos que mantenham as Bases de Conhecimento Organizacionais.

KIDD (1994) analisa que, normalmente, os Processos Intensivos em Conhecimento envolvem um custo muito alto, devido à forte dependência do conhecimento dos participantes envolvidos como ocorre, por exemplo, em processos de Desenvolvimento de Software ou em processos da administração pública.

Processos de Negócio tradicionais são caracterizados pelas estruturas predefinidas de processos, e suas tarefas repetidas realizadas com base nos modelos de processo básico, contendo informações, tarefas e papéis de usuários. Processos Intensivos em Conhecimento são apenas parcialmente mapeados devido às decisões imprevisíveis e tarefas guiadas por criatividade, onde, tipicamente, os fluxos de conhecimento e a troca de conhecimento entre os participantes são necessários para alcançar a execução com sucesso destes processos, sendo esta realizada de maneira tipicamente *ad hoc*.

Apesar do papel fundamental dos Processos Intensivos em Conhecimento para a Gestão do Conhecimento, HOFFMANN (2002) afirma que o apoio da TI aos Processos Intensivos em Conhecimento não é ainda adequado, uma vez que estes são fortemente baseados na socialização e na troca informal de conhecimento.

Com base nesta afirmação, assume-se que há conhecimento tácito intrínseco no conteúdo textual gerado pela troca informal de conhecimento, através de Ferramentas Colaborativas. Por esse motivo, esta pesquisa objetiva apresentar um novo método para descoberta de Processos Intensivos em Conhecimento, a partir de ferramentas colaborativas, como e-mails. Para isso, será realizado, a seguir, um estudo, entre as propostas existentes, para a representação de Processos Intensivos em Conhecimento.

2.4.1 Formas de Representação

Para PAPAVALASSILIOU *et al.* (2002), o hiato entre a representação de Processos de Negócio e seus atributos de conhecimento tem se mostrado crítico para organizações, considerando que o conhecimento é fator crucial para o sucesso do negócio.

GRONAU & WEBER (2003) ressaltam que a ocorrência do conhecimento, através da sua transferência entre agentes, não é suficientemente modelada em ferramentas convencionais de modelagem de Processos de Negócio. Por esse motivo, para a representação de Processos Intensivos em Conhecimento, devem ser considerados os seguintes pontos:

- Integrar os modelos de representação de processo e de conhecimento;
- Possibilitar uma representação do modelo de negócio organizacional considerando os processos de negócio e suas atividades de conhecimento;
- Apoiar a Gestão do Conhecimento Orientada a Processos de Negócio.

Visando atender a estas questões, existem, na literatura, algumas iniciativas que têm como principal objetivo possibilitar a representação de aspectos de conhecimento integrados aos Processos de Negócio. Neste estudo, serão brevemente listadas as três principais propostas.

PAPAVALASSILIOU (2002) apresenta o BPK (*Business Process Knowledge*) que foca no formalismo do detalhamento dos processos, provendo uma série de conceitos instanciáveis que garantem a dinamicidade da representação.

ABECKER *et al.* (2001) propõem o método DECOR, cujo foco é voltado para a representação de Processos Intensivos em Conhecimento através da criação de diagramas que buscam estruturar aos Processos de Negócio, o contexto dinâmico, as informações contextualizadas e representações das memórias embutidas ao processo produtivo.

GRONAU (2004) propõe o KMDL (*Knowledge Management Description Language*), cujo objetivo trata da criação, utilização e necessidade do conhecimento ao longo dos Processos de Negócio, oferecendo recursos para sua formalização, através da utilização de características específicas de conhecimento.

Como se torna evidente, através das propostas analisadas, as principais abordagens existentes mantêm seu foco na representação de Processos Intensivos em Conhecimento utilizando recursos existentes da Modelagem de Conhecimento.

No entanto, GRONAU & WEBER (2003) afirmam que, mesmo com essas iniciativas, ainda não há linguagens ou ferramentas capazes de separar informação tácita de explícita. Sendo assim, conclui que estas abordagens não permitem visualizar, de maneira adequada, as decisões, ações e métricas envolvidas na sequência de atividades executadas nos processos.

Por essa razão, nesta pesquisa, para fins de representação dos Processos Intensivos em Conhecimento a serem descobertos pelo método proposto, serão adotados Mapas Mentais como recurso gráfico de representação, uma vez que são conhecidos, flexíveis e permitem representar, de forma clara, objetiva e intuitiva, as informações relacionadas a estes processos.

Contudo, para que Processos Intensivos em Conhecimento possam ser descobertos pelo método proposto nesta pesquisa, é fundamental que seja realizada uma análise detalhada dos Elementos que os compõem. Esta análise permitirá compreender quais são, de fato, as informações relevantes necessárias a serem descobertas pela solução proposta, de forma que atenda a seu objetivo. O estudo a seguir trata de identificar e conceituar os principais elementos existentes em um Processo Intensivo em Conhecimento.

2.4.2 Elementos

Segundo GRONAU (2004), para que seja possível representar Processos Intensivos em Conhecimento, é necessário que os seguintes requisitos sejam explicitados:

- **Portador do Conhecimento**, uma vez que o conhecimento é limitado aos indivíduos;
- **Descritor do Conhecimento**, uma vez que o conhecimento, por si só, não pode ser codificado, a descrição do domínio do conhecimento faz-se necessária;
- **Classe de Competência**, uma vez que o conhecimento necessário ou disponível de um indivíduo pode atender a um conjunto arbitrário de domínios de conhecimento comuns definidos.
- **Objeto de Conhecimento**, uma vez que o conhecimento modelado por um indivíduo é referente a uma seção do domínio;
- **Demanda de Conhecimento**, uma vez que o conhecimento pode ser descoberto ou disponibilizado;
- **Objeto de Informação**, uma vez que o conhecimento explícito não é limitado a pessoas;
- **Fluxo de Conhecimento**, uma vez que o conhecimento pode ser externalizado, internalizado, socializado ou combinado.

Visando a atender aos requisitos necessários apresentados por GRONAU (2004), no escopo desta pesquisa, serão adotados os seguintes elementos para a descoberta de um Processo Intensivo em Conhecimento: Processos, Participantes, Interações Sociais, Atividades, Itens de Conhecimento e Artefatos de Conhecimento.

Processo é o Descritor do Conhecimento e deverá ser aplicado para descrever os limites e conteúdo do domínio de conhecimento, não tratando da sua codificação. O método apresentado nesta pesquisa assume que este elemento, ou seja, o Processo Intensivo em Conhecimento que se deseja descobrir é conhecido, de forma que seus demais elementos possam, a partir dele, ser descobertos.

Participante é o Portador do Conhecimento e deverá conter, portanto, o conhecimento necessário para a realização de suas tarefas. Por esse motivo, a ele serão associadas todas as Atividades, Itens de Conhecimento e Artefatos de Conhecimento por ele realizados.

Interação Social é o Fluxo de Conhecimento propriamente dito e deverá conter o relacionamento entre os participantes, ou seja, a troca de conhecimento existente. Em cada Interação Social, deverão ser atribuídas as Atividades, Itens de Conhecimento e Artefatos de Conhecimento relacionados. Além disso, deverá ser exibida a frequência de participação de cada Participante durante as Interações Sociais realizadas.

Atividade é a Demanda de Conhecimento, uma vez que, a partir da sua execução, é possível descobrir o conhecimento existente. As Atividades refletem a estrutura básica dos modelos de Processo de Negócio e a sua sequência de execução determina a estrutura temporal do processo.

Item de Conhecimento é o Objeto de Conhecimento e, portanto, deverá descrever o conhecimento dos Participantes envolvido na execução do processo e suas atividades.

Artefato de Conhecimento é o Objeto de Informação e, portanto, deverá descrever todo e qualquer conhecimento explícito gerado a partir da execução do processo ao qual pertence.

Com base na relação apresentada de elementos necessários para a descoberta e representação de Processos Intensivos em Conhecimento, no escopo desta pesquisa, serão considerados todos estes elementos como informações relevantes para a descoberta destes processos.

Portando, o método proposto terá como objetivo extrair, a partir de uma amostra de e-mails relacionada ao Processo Intensivo em Conhecimento que se deseja descobrir, todas as informações relevantes a sua representação, ou seja, Processo, Participantes e suas Interações Sociais, Atividades, Itens de Conhecimento e Artefatos de Conhecimento.

2.5 Conclusão

Numa perspectiva de Sistemas de Informação, o Conhecimento é considerado orientado a propósito e à resolução de problemas, composto por informações contextualizadas e interligadas e limitado aos modelos internos dos indivíduos. A criação do conhecimento organizacional deve ser compreendida como um processo que amplifica organizacionalmente o conhecimento criado pelos indivíduos e o consolida como parte da sua rede de conhecimentos.

A Gestão do Conhecimento apresenta-se como a área de pesquisa que visa adquirir, processar e criar conhecimento organizacional acessível, de forma sistemática, visando prover as melhores decisões e estar mais bem preparado para o futuro, através da identificação, desenvolvimento e provisão do conhecimento relevante ao sucesso da organização.

A Gestão do Conhecimento Centrada em Processos visa promover, motivar, nortear e guiar o processo de conhecer, descartando as tentativas de capturar e distribuir conhecimento, compreendendo que a Gestão do Conhecimento trata de um processo de comunicação social, o qual deve ser apoiado por ferramentas colaborativas.

Esta abordagem defende, inclusive, a existência de conhecimento a partir das atividades executadas por Agentes em Processos de Negócio. Por essa razão, a Gestão do Conhecimento apresenta-se como possível forma de apoio à descoberta de conhecimento através dos Processos de Negócio, provendo os métodos e ferramentas necessários. Para atender a este objetivo, foi criada uma subárea de pesquisa denominada Gestão do Conhecimento Orientada a Processos de Negócio.

Neste contexto, é possível denotar a relevância dos Processos Intensivos em Conhecimento, também conhecidos como Processos de Negócio Intensivos em Conhecimento, uma vez que identificar, modelar, analisar e otimizar estes processos deve ser o principal objetivo, em longo prazo, da Gestão do Conhecimento Orientada a Processos.

Os Processos Intensivos em Conhecimento envolvem um custo muito alto, devido à forte dependência do conhecimento por parte dos participantes envolvidos. Por essa razão, devem ser o cerne dos processos organizacionais, de forma a produzir ou criar novos conhecimentos que mantenham as Bases de Conhecimento Organizacionais.

Para isso, existem diversas propostas na literatura que objetivam apoiar a esta necessidade. No entanto, ainda não há linguagens ou ferramentas capazes de separar informação tácita de explícita, de forma a permitir a visualização de decisões, ações e métricas envolvidas na sequência de atividades executadas nos processos.

Por essa razão, foi definida, no contexto desta pesquisa, a utilização de Mapas Mentais como recurso gráfico de representação, uma vez que permitem representar, de forma clara, objetiva e intuitiva, os Processos Intensivos em Conhecimento.

Da mesma forma, serão consideradas, no escopo desta pesquisa, informações relevantes à descoberta de um Processo Intensivo em Conhecimento, os seguintes elementos: Processos, Participantes e suas Interações Sociais, Atividades, Itens de Conhecimento e Artefatos de Conhecimento.

Capítulo 3 – O Método *Story Mining*

Neste capítulo será abordado o método *Story Mining*, uma abordagem que visa auxiliar a Elicitação de Processos de Negócio através da exploração de narrativas colaborativas (*Group Storytelling*) associadas a técnicas de Mineração de Texto, apresentando os principais conceitos e técnicas adaptadas para atender a esta pesquisa.

3.1 Introdução e Trabalhos Relacionados

O método proposto pela presente pesquisa visa à descoberta de Processos Intensivos em Conhecimento a partir da extração de informações relevantes contidas em e-mails, onde a troca informal de conhecimento é uma realidade.

Um dos elementos principais para a descoberta e, conseqüente, representação destes processos são as Atividades, visto que, através da execução destas, é possível descobrir o conhecimento existente. Além disso, as Atividades refletem a estrutura básica dos modelos de Processo de Negócio e a sua seqüência de execução determina a estrutura temporal do processo.

No entanto, apesar de sua relevância, a descoberta de Atividades torna-se um dos maiores desafios do método proposto por esta pesquisa, uma vez que a análise proposta é realizada com base em uma amostra de e-mails selecionada, onde as entidades atômicas referentes às Atividades não podem ser diretamente extraídas. Por esta razão, torna-se necessária a utilização de uma solução capaz de extrair Atividades, a partir do conteúdo textual informal, ou seja, do texto não estruturado contido nas trocas de e-mails existentes.

Existem, na literatura, algumas propostas relacionadas a esta necessidade. VAN DER AALST (2005) apresenta o conceito da Mineração de Processos (*Process Mining*), propondo um arcabouço de mineração de processos, o ProM (*Process Mining Workbench*). O seu principal objetivo consiste em extrair informações organizacionais sobre processos a partir de *logs* de eventos gerados por Sistemas de Gestão de Workflows, como sistemas de ERP, e gerar, com base nos dados extraídos, um modelo representativo do processo utilizando notações como Redes de PETRI e EPC.

Apesar da relevância da proposta de VAN DER AALST (2005), a solução baseia-se na extração de informações, a partir de *logs* de eventos gerados por sistemas de *workflow*, ou seja, na mineração de dados estruturados, não sendo aplicável, portanto, às necessidades desta pesquisa.

VAN DER AALST & NIKOLOV (2007) propõem o *EmailAnalyzer*. Esta adaptação é responsável por extrair, a partir de e-mails, as informações necessárias para a estruturação de *logs* de eventos, de forma que possam ser aplicados ao ProM. O principal objetivo desta adaptação trata da análise de Redes Sociais que permita a descoberta de interações sociais existentes nos processos de negócio.

Seu foco é totalmente voltado à descoberta de informações básicas contidas nos e-mails, como Data de Envio, Remetente e Destinatário, não considerando as atividades e o papel dos envolvidos diretamente com a sua execução. Logo, esta proposta não atende à descoberta de Atividades que envolvam decisões e que possuam um fator humano externo aos sistemas e, portanto, não atende ao escopo desta pesquisa.

UDOH (2007) propõe uma abordagem baseada em técnicas de mineração de dados voltadas à análise de e-mails, capaz de analisar e minerar informações contidas em servidores de e-mail POP3, provendo um apoio computacional à tomada de decisões. Apesar de a solução proposta ser capaz de extrair, classificar e agrupar dados que auxiliem a tomada de decisão, esta não oferece uma solução para a descoberta de Atividades a partir da mineração de e-mails.

GONÇALVES (2010) apresenta o *Story Mining*, um método para a Elicitação de Processos de Negócio através da exploração de narrativas colaborativas (*Group Storytelling*) associadas a técnicas de Mineração de Texto. Nesta abordagem, GONÇALVES (2010) defende que é possível descobrir Atividades Candidatas de processos de negócio, a partir da análise de textos não estruturados no formato de narrativas, utilizando técnicas de mineração de texto.

Apesar de seu foco ser voltado a narrativas colaborativas, esta solução apresenta-se como análoga às necessidades desta pesquisa, uma vez que se deseja descobrir Atividades a partir de textos não estruturados contidos nos e-mails selecionados. Portanto, na solução proposta pela presente pesquisa, será adotado o método *Story Mining* para a descoberta automática das Atividades relacionadas aos Processos Intensivos em Conhecimento analisados.

Vale ressaltar que, no escopo da presente pesquisa, serão necessárias adaptações a alguns componentes do método *Story Mining* que permitirão a integração deste método à solução proposta por esta pesquisa. Estas adaptações serão descritas detalhadamente no próximo capítulo.

Neste capítulo, serão detalhados os principais conceitos necessários à perfeita compreensão do método *Story Mining* e apresentadas as fases do método executadas de forma a obter os resultados esperados nesta pesquisa.

3.2 Mineração de Textos

Analogamente à mineração tradicional, onde se busca, em meio a pedras e terra, encontrar metais preciosos, HEARST (1999) afirma que a Mineração de Texto consiste na extração de informação ou conhecimento previamente desconhecido a partir da análise de textos. Trata-se de uma área de pesquisa relativamente recente, se comparada a áreas tradicionais da Ciência da Computação, onde o enfoque principal encontra-se na análise de textos não estruturados, diferentemente da Mineração de Dados, onde o principal objetivo consiste na extração de padrões a partir de formas estruturadas de dados, como Bancos de Dados Relacionais.

SCHWARTZ & HEARST (2003) conceituam a Mineração de Texto como um processo responsável por extrair automaticamente, a partir de diversos recursos de texto, novas informações, auxiliando a descoberta de novos fatos, novas hipóteses e conclusões previamente desconhecidas.

FELDMAN & SANGER (2007) definem a Mineração de Textos como o Processo Intensivo em Conhecimento onde um indivíduo interage com uma coleção de documentos, através da utilização de uma série de ferramentas de análise.

HOTHOTH *et al.* (2003) afirmam que a Mineração de Textos consiste em um método verdadeiramente interdisciplinar, uma vez que aplica conceitos e técnicas provenientes de diversas áreas do conhecimento como: Recuperação da Informação, Extração da Informação e Processamento da Linguagem Natural. Para fins de conceituação, segue uma breve descrição destas áreas:

- **Recuperação da Informação:** lida com a representação, armazenamento, organização e o acesso a itens de informação, visando fornecer ao usuário acesso fácil à informação desejada (BAEZA YATES & RIBEIRO NETO, 1999).
- **Extração da Informação:** consiste em uma forma limitada de compreender textos onde já se conhece o tipo de informação que se deseja extrair, ou seja, na extração de informação relevante a partir de documentos (HOTHOTH *et al.*, 2005).
- **Processamento da Linguagem Natural:** consiste em prover meios de alcançar um maior grau de compreensão da linguagem natural, permitindo compreender a estrutura gramatical e semântica da linguagem através do uso de computadores (KODRATOFF, 1999).

Das áreas de conhecimento citadas, no contexto de aplicação do método *Story Mining*, serão utilizadas técnicas relacionadas à Extração da Informação e Processamento da Linguagem Natural, as quais serão apresentadas mais adiante. A seguir, serão apresentados os principais conceitos relacionados à Mineração de Texto.

3.2.1 Conceitos Básicos

Antes de transcorrer sobre as técnicas adotadas ao longo da execução das fases do método *Story Mining*, é necessário que sejam apresentados alguns conceitos básicos que permitirão a compreensão necessária do tema, sendo estes: Documentos, Coleções de Documentos, Modelos Representativos de Documentos e Domínio e Conhecimento Externo.

3.2.1.1 Documento

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

Embora, em alguns pontos desta pesquisa, sejam feitas menções a documentos de texto não estruturados, os documentos podem ser classificados, segundo a sua estrutura, em duas categorias: semiestruturados e fracamente estruturados. São denominados semiestruturados, todos os Documentos baseados em texto que possuam uma estrutura interna e modelos a serem obedecidos, como páginas da Web, arquivos PDF, XML e DOC. Em contrapartida, são classificados como fracamente estruturados, todos os Documentos de texto livre, compostos apenas de texto e com pouca informação adicional em sua estrutura.

3.2.1.2 Modelo Representativo do Documento

GONÇALVES (2010) define que para tornar possível a extração de conhecimento, o documento deve ser definido em relação aos elementos que o compõem e quais relações e sentidos estes elementos possuem entre si. Este conjunto de elementos é denominado Modelo Representativo do Documento, sendo responsável por descrever os diferentes elementos que um tipo abstrato de documento possui.

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. Quando aplicado a um documento fracamente estruturado, o Modelo Representativo do Documento será composto por elementos de texto, como palavras, frases e parágrafos.

A vasta gama de possíveis modelos representativos de documentos apresenta uma série de dificuldades, principalmente, com relação à quantidade de elementos que pode possuir. Por este motivo, embora exista uma diversidade de elementos a serem utilizados para representar um documento, no escopo do *Story Mining* foram utilizados quatro tipos básicos de elementos: Caracteres, Palavras, Termos e Conceitos.

- **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;
- **Palavras:** consistem no nível mais básico, em termos de riqueza semântica, presentes nos Documentos, diferenciando-se dos caracteres por possuírem um significado próprio. Existem diversas técnicas voltadas para este tipo de elemento, como filtragem de *stop words* (Palavras de Parada), que permite a eliminação de palavras sem sentido semântico e a frequência de palavras, que permite identificar quais palavras são mais representativas dentre todos os documentos;
- **Termos:** consistem em palavras ou frases de várias palavras identificadas, através de padrões morfossintáticos, e selecionadas diretamente do corpo de um documento por meio de um método de extração de termos, oferecendo um nível semântico ainda maior que as palavras. O seu objetivo é a representação reduzida do documento, com uma riqueza semântica maior do que a simples extração de palavras (FELDMAN *et al.*, 1998).

- **Conceitos:** consistem em elementos de uma riqueza semântica mais alta do que os termos, representando abstrações que podem ser refletidas em fragmentos de texto. 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, uma ontologia ou até mesmo pelo *feedback* com o usuário (BLAKE & PRATT, 2001).

3.2.1.3 Coleção de Documentos

FELDMAN & SANGER (2007) definem a Coleção de Documentos como sendo qualquer grupo de documentos baseados em texto. Por esta razão, trata-se de um conceito fundamental à Mineração de Textos, uma vez que, para que sejam extraídos relações e padrões relevantes sobre um assunto específico, é fundamental a aplicação de grandes coleções de documentos.

SANTOS *et al.* (2009) afirmam que há diversas técnicas que visam as coleções para mapear termos e conceitos presentes em seus documentos e extrair conhecimento a partir dos padrões encontrados. Com o advento da Web, houve uma forte expansão deste campo de pesquisa, sendo utilizada para objetivos mais complexos como a descoberta de relacionamentos em potencial dentro de redes sociais.

3.2.1.4 Domínio 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, como Aviação, Economia, entre outros, permitem trabalhar eficientemente com conceitos presentes em um texto, ampliando as possibilidades de aplicações da área. 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 extraíndo conhecimento de melhor qualidade.

De forma geral, estas fontes se traduzem de alguma forma estruturada até formas mais complexas como ontologias de domínio. Na prática, a principal função destas fontes é facilitar a descoberta do conhecimento, permitindo identificar parcelas de texto relevantes dentre todos os documentos disponíveis.

3.2.2 Processo de Mineração de Textos

Segundo GONÇALVES (2010), o Processo Intensivo em Conhecimento de Mineração de Textos utiliza fontes de texto como entrada, como documentos e coleções de documentos e gera uma série de saídas, geralmente envolvendo alguma forma estruturada de conhecimento e informação útil para o usuário, como padrões encontrados, visualizações, tendências, gráficos, entre outros. Para atender a este objetivo, em uma visão alto nível, este processo pode ser dividido em três fases: Pré-Processamento, Operações de Mineração e Pós-Processamento, conforme ilustra a Figura 3.1:



Figura 3.1 – Visão Macro do Processo de Mineração de Textos (GONÇALVES, 2010)

3.2.2.1 Pré-Processamento

Esta fase envolve a aplicação de diversas rotinas, algoritmos e operações destinadas a preparar o dado original, em forma bruta, para que seja processado pelas fases posteriores. Sendo assim, o enfoque desta fase consiste na explicitação ou geração de modelos representativos dos documentos de entrada. Por essa razão, é tão comum o uso de técnicas de Processamento da Linguagem Natural nesta fase, uma vez que possui um papel fundamental na criação de uma estrutura a partir dos elementos básicos de texto.

Algumas técnicas de Processamento da Linguagem Natural utilizadas nesta fase são:

- **Filtragem por elementos básicos e stopwords:** consiste em um conjunto de técnicas responsáveis por eliminar, a partir do texto original, elementos considerados irrelevantes, como palavras frequentes ou impertinentes a uma aplicação (DOLAMIC & SAVOY, 2009).
- **Ocorrência de Palavras:** consiste em um dos padrões mais básicos de extrair a partir de documentos. A frequência de palavras é uma técnica muito utilizada para a indexação e classificação de documentos e para a descoberta automática de palavras-chaves representativas de um domínio (FELDMAN & SANGER, 2007).
- **Resolução de Anáforas e Correferência:** consiste em um conjunto de algoritmos responsáveis por resolver as anáforas existentes em documentos de texto. Anáfora é o fenômeno linguístico onde um trecho do discurso remete a um referente, sendo esta uma palavra já mencionada, que ainda será mencionada ou externa ao discurso (CHAVES & RINO, 2008).
- **Part-Of-Speech Tagging (POS-Tagging):** consiste em um conjunto de técnicas de classificação de palavras, onde o contexto de uma palavra reflete o papel exercido por ela em uma frase. O conjunto (*tagset*) mais comum de etiquetas de classificação (*tags*) é: Artigo, Substantivo, Verbo, Adjetivo, Preposição, Numeral e Nome Próprio. Estas técnicas utilizam usam corpora (*sing. corpus*) de documentos para treinamento ou um conjunto de regras para guiar a tarefa de classificar automaticamente as palavras de um texto não classificado. Uma das técnicas mais comuns de POS-Tagging é o N-GramTagger, um algoritmo que possibilita a classificação morfossintática de palavras, a partir do treinamento de corpora de textos (BIRD *et al.*, 2009).

- **Parsing Sintático:** consiste em um conjunto de técnicas responsáveis por realizar análises sintáticas aprofundadas nas frases do texto analisado, com base em uma teoria gramatical (FELDMAN & SANGER, 2007).
- **Shallow Parsing:** consiste em um conjunto de algoritmos que propõe a simplificação da análise sintática em prol de uma maior velocidade de processamento e robustez das técnicas. Dessa forma, estes algoritmos propõem a extração de simples sintagmas nominais e verbais, criando uma estrutura inicial a partir da qual as classificações mais complexas serão formadas (SANG & BUCHHOLZ, 2000).

3.2.2.2 Operações de Mineração

Segundo GONÇALVES (2010), as operações de mineração consistem no núcleo fundamental da aplicação de mineração de textos, tendo como objetivo a extração de padrões, identificação de tendências e aquisição de novos conhecimentos e informações previamente desconhecidos, a partir do texto pré-processado. Nesta fase, é fundamental a aplicação de técnicas de Extração de Informação. Embora estas técnicas possuam objetivos distintos da área de Mineração de Texto, elas podem ser utilizadas, como parte de um contexto maior e em combinação com outras áreas, como aplicado no método *Story Mining*.

3.2.2.3 Pós-Processamento

GONÇALVES (2010) define que 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.

3.3 O Método *Story Mining*

O método *Story Mining* proposto por GONÇALVES, SANTORO, BAIÃO (2009) consiste em uma abordagem para uma nova forma de elicitação de Processos de Negócio, oferecendo uma melhoria à área de Elicitação de Processos de Negócio e, conseqüentemente, melhor qualidade para a Modelagem de Processos de Negócio. Para atender a este objetivo, foram definidos cinco requisitos principais:

- Utilizar narrativas, como forma de expressão direta de extração de conhecimento dos participantes do processo;
- Possuir um enfoque na característica colaborativa da elicitação;
- Permitir que as pessoas envolvidas no processo de elicitação expressem seus pontos de vista de forma mais livre possível;
- Minimizar os possíveis efeitos negativos da ação do analista na tarefa de elicitação e modelagem como um todo;
- 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.

Por atender a estes objetivos, GONÇALVES (2010) subdivide o método em três fases principais: Construção Coletiva de Histórias, Mineração de Texto e Construção de Modelo Formal, conforme ilustradas na Figura 3.2:

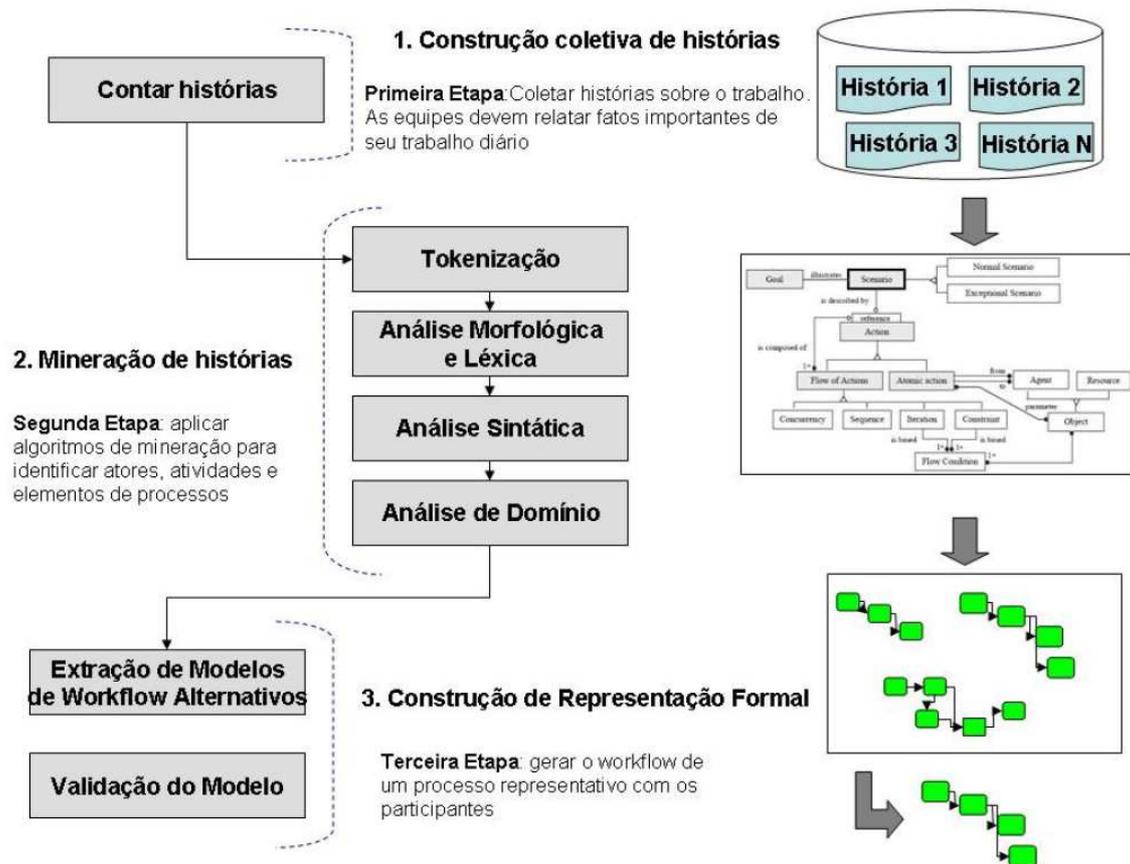


Figura 3.2 – Fases do Método *Story Mining* (GONÇALVES, 2010)

A primeira fase do método, Construção Coletiva de Histórias, consiste na narrativa colaborativa de fatos concretos pelos participantes e seu posterior armazenamento em um repositório de histórias. Esta ideia é baseada na proposta de SANTORO *et al.* (2008) que propõem, de forma preliminar, um método para elicitación de processos de negócio baseado em *Group Storytelling*, a partir da qual os atores do processo descrevem colaborativamente sua forma de agir através de histórias.

No entanto, GONÇALVES *et al.* (2009) afirmam que o conteúdo informal das narrativas e sua descrição em texto livre proporcionam uma possível dificuldade adicional ao Analista de Processos, devido à quantidade de informações a ser lida e interpretada. Este risco traz a necessidade de técnicas que provejam a visualização, de forma mais clara, do conhecimento presente no repositório.

No método *Story Mining*, é proposto que esta tarefa seja realizada através da aplicação de técnicas de Mineração de Texto, Extração de Informação e Processamento da Linguagem Natural, visando mitigar este problema.

Por esse motivo, o método *Story Mining* foi desenvolvido, de modo que a extração automática de atividades e a construção de abstrações do processo fossem apoiadas por um processo de mineração de textos, sendo este executado por uma ferramenta de gerência de workflows. De forma a atender às necessidades de conceituação apresentadas por esta pesquisa, serão detalhadas nesta seção apenas as funcionalidades do método *Story Mining* aplicadas com o fim de atender à solução proposta neste trabalho.

3.3.1 Fases da Mineração de Histórias

Uma vez definido o objetivo, serão apresentadas as técnicas de Processamento de Linguagem Natural utilizadas para a identificação dos elementos básicos de processos presentes na narrativa e suas relações lógico-temporais. Para isso, a aplicação dos algoritmos se divide em quatro etapas: Tokenização, Análise Léxica e Morfológica, Análise Sintática e Análise de Domínio.

3.3.1.1 Tokenização

Nesta etapa, GONÇALVES (2010) propõe a extração dos componentes básicos do texto das histórias, 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 original para o processamento computacional.

3.3.1.2 Análise Léxica e Morfológica

Nesta etapa, GONÇALVES (2010) defende que a lista gerada pela etapa anterior, composta por palavras, pontuação e outros elementos de texto, sejam analisados de modo a aprofundar o processo de estruturação do texto, ampliando a estrutura básica de palavras para uma classificação morfológica e adicionando informação gramatical a cada elemento da lista.

3.3.1.3 Análise Sintática

Nesta etapa, GONÇALVES (2010) propõe o refinamento do texto através do agrupamento de palavras conforme suas respectivas classificações morfológicas, de forma que cada grupo possua informação morfossintática suficiente para a identificação de padrões. Estes grupos são classificados como termos de Sintagmas Nominais e Sintagmas Verbais, Frases e outros tipos de estruturas mais complexas. A Figura 3.4 ilustra um exemplo de texto gerado a partir da 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 3.3 – Texto Gerado a partir da Análise Sintática (GONÇALVES, 2010)

3.3.1.4 Análise de Domínio

Nesta última etapa, GONÇALVES (2010) defende a extração de elementos do *template*, incluindo seus relacionamentos e o seu posterior preenchimento de cenário, visando obter, como produto final estruturado, os candidatos a elementos de processo de negócio. Para isso, propõe que a estrutura gerada seja processada, de forma que sejam identificados padrões gramaticais que possam se adequar ao template utilizado como guia para a descoberta de candidatos a elementos de processos, considerando, por exemplo, Sintagmas Nominais como candidatos a Atores e Sintagmas Verbais como candidatos a ações.

A Figura 3.4 ilustra o resultado final gerado pela última etapa da fase de Mineração de Histórias proposta:

```
<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 3.4 – Exemplo de Resultado Gerado pela Fase de Mineração de Histórias (GONÇALVES, 2010)

3.3.2 Solução Proposta pelo *Story Mining*

O processo de mineração de textos pode ser visto como um workflow científico. Devido à grande variedade de técnicas, algoritmos e programas de Mineração de Texto de Processamento da Linguagem Natural utilizados em cada passo do método proposto, GONÇALVES (2010) propõe a adoção de um Sistema de Gerenciamento de Workflows Científicos (SGWfC).

CAVALCANTI, TARGINO, BAIÃO *et al.* (2005) definem um workflow científico como uma combinação de dados e sequência de programas que, uma vez executados, caracterizam um experimento através de suas entradas e saídas. Um SGWfC permite a execução e análise comparativa entre diversas instâncias do workflow, através da utilização de diversas técnicas, programas e valores de argumentos, permitindo, até mesmo, fontes de conhecimento e entradas para o método como um todo.

Para a implementação das técnicas de Mineração de Texto, GONÇALVES (2010) adota o Sistema de Gerenciamento de Workflows Científicos *VisTrails*, devido às suas funcionalidades avançadas para visualização e comparação de resultado, além de sua extensibilidade nativa para a integração com aplicações externas, por meio de módulos extras desenvolvidos, como classes, em linguagem Python e inseridos no *VisTrails* através de bibliotecas extras de módulos para workflows científicos. A Figura 3.5 ilustra o workflow de mineração de texto proposto por GONÇALVES (2010) como representado pela interface do SGWfC *VisTrails*:

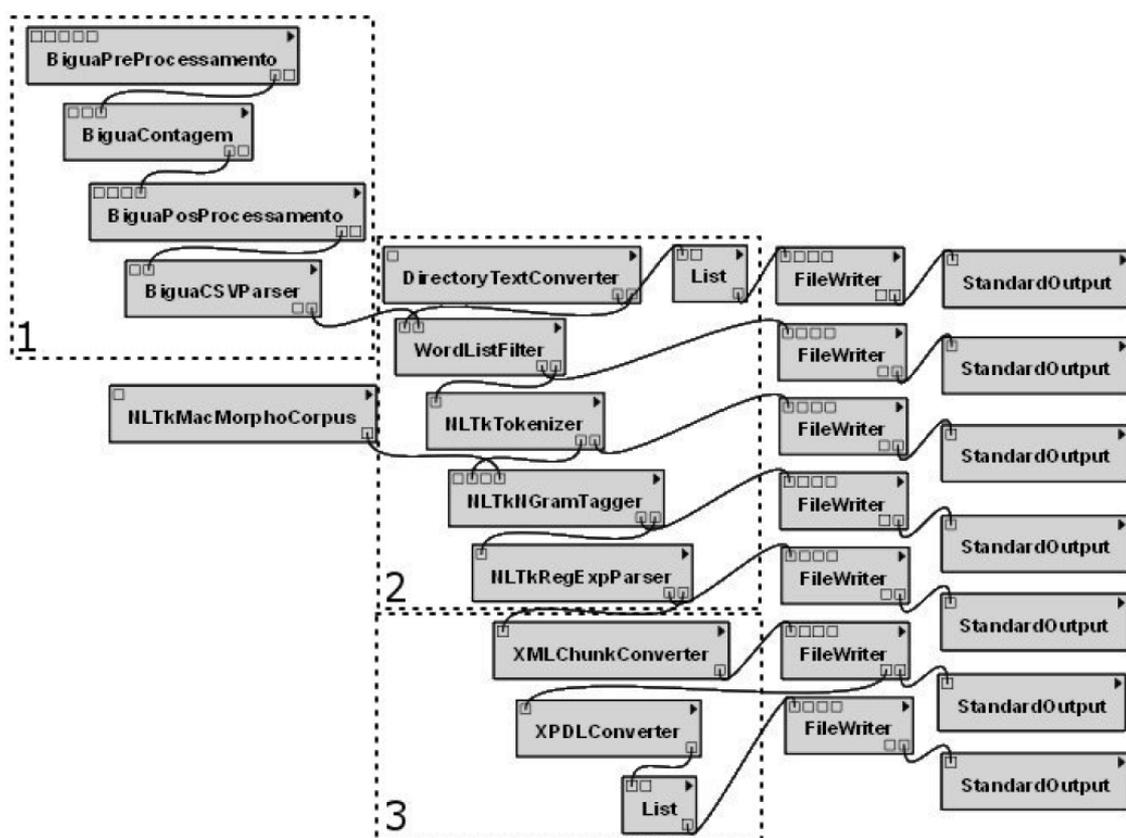


Figura 3.5 – Workflow de Mineração de Texto do *Story Mining* (GONÇALVES, 2010)

GONÇALVES (2010) divide o workflow proposto pelo método *Story Mining* em três fases distintas, conforme destacado na figura 3.6. Na primeira fase, são aplicadas técnicas de cálculo de frequência de palavras e radicalização de termos aos documentos relacionados à história analisada.

Na segunda fase, GONÇALVES (2010) propõe a extração dos eventos da história, os quais são convertidos para texto puro e aplicados a uma sequência de técnicas de pré-processamento, desde a filtragem de frases relevantes até a análise sintática e a estruturação final do texto. Na terceira fase, GONÇALVES (2010) propõe a conversão do texto estruturado em formatos específicos para processos de negócio.

Vale ressaltar que os módulos *FileWriter* e *StandardOutput* vinculados às classes de execução do workflow, têm como finalidade a geração de arquivos de *log* parciais do processo. Estes logs proveem as informações necessárias para a análise de possíveis erros e desempenho do método, tal como a avaliação das técnicas adotadas. No entanto, estes módulos não são considerados como parte do processo de mineração de textos propriamente dito, sendo apenas utilizados para escritas de arquivos em disco.

Nesta seção serão descritas detalhadamente as técnicas adotadas em cada fase da solução proposta.

3.3.2.1 Processamento dos Documentos

Nesta fase inicial, para a sua operacionalização, foram criados os módulos *BiguaPreProcessamento*, *BiguaContagem* e *BiguaPosProcessamento*.

O Biguá consiste em uma plataforma de desenvolvimento de aplicações de mineração de texto, com suporte aos idiomas português e inglês. Para o método Story Mining, GONÇALVES (2010) adota a versão Workflow Biguá (OLIVEIRA, 2008) para a implementação do pré-processamento e da aplicação de técnicas de cálculo de frequência de palavras nos documentos relacionados às histórias.

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

De modo a evitar a ambiguidade entre palavras com significados semelhantes, GONÇALVES (2010) aplica uma técnica adicional, denominada *RSLP Stemmer*, visando a redução das palavras aos seus radicais, refinando a filtragem de palavras de domínio do método.

Finalmente, nesta fase, é gerado 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.

3.3.2.2 Processamento de Eventos

Primeiramente, GONÇALVES (2010) propõe a criação do módulo *DirectoryTextConverter*, responsável por converter o conteúdo dos eventos da história analisada em um arquivo de texto puro. Após isso, é criado o módulo *WordListFilter*, visando a filtragem de frases a partir da lista de palavras extraída na fase anterior.

Em seguida, GONÇALVES (2010) propõe a utilização de diversas técnicas de mineração de texto, baseado em Processamento de Linguagem Natural. Para isso, adota o framework NLTK (*Natural Language Toolkit*), devido à vasta gama de algoritmos de Parsing Sintático, *Shallow Parsing* e Tokenização disponíveis, além de oferecer diversos corpora de documentos etiquetados em múltiplos idiomas.

Uma vez definido framework a ser utilizado, GONÇALVES (2010) aplica a técnica de Tokenização, através do módulo *NLTKTokenizer*, responsável por gerar uma lista de palavras e pontuação a partir do texto filtrado.

Concluída esta etapa, é aplicada a técnica de *Trigram Tagger*, através do módulo *NLTKNGramTagger*, para a criação da estrutura gramatical a partir de palavras não classificadas.

Para isso, GONÇALVES (2010) utiliza um corpus de treinamento MAC-MORPHO (ALUISIO *et al.*, 2004), através do módulo *NLTKMacMorphoCorpus*, para classificar cada palavra. Esta classificação inicial é refinada pelo módulo *NLTKRegExpParser*, que aplica a técnica de *Shallow Parsing*, a partir de uma gramática de expressões regulares. Dessa forma, é possível obter uma estrutura mais refinada e de diversos níveis, que servirá como base para a última fase do workflow de mineração de texto.

3.3.2.3 Extração de Informação

Nesta última fase, GONÇALVES (2010) propõe a extração e conversão dos elementos de processo candidatos em duas formas específicas, sendo estes elementos: Atores, Atividades e os Parâmetros envolvidos, uma vez que estes constituem os principais elementos de um processo de negócio convencional.

Um formato consiste em um arquivo XPDL, gerado pelo módulo *XPDLConverter*, uma vez que o método *Story Mining* propõe a elicitación de processos de negócio. Portanto, através da geração de um arquivo no formato XPDL torna-se possível a sua representação através de diversas ferramentas de modelagem de processos de negócio.

O segundo formato, gerado pelo módulo *XMLChunkConverter*, consiste em uma estrutura XML, onde são estruturadas todas as atividades, seguidas de seus respectivos atores e parâmetros.

No contexto desta pesquisa, acredita-se que o resultado final gerado pelo método *Story Mining*, onde são estruturadas todas as atividades candidatas descobertas em um arquivo XML, seja uma solução que atende a uma das principais necessidades desta pesquisa. Por essa razão, o método *Story Mining* foi selecionado para integrar a Proposta de Solução a ser detalhada a seguir.

Capítulo 4 – Proposta de Solução

Este capítulo tem como objetivo apresentar detalhadamente o método para descoberta de Processos Intensivos em Conhecimento proposto por esta pesquisa, descrevendo os passos necessários para obtenção do resultado esperado.

4.1 Requisitos da Proposta

Conforme apresentado nos capítulos anteriores, esta proposta de solução está baseada nos seguintes requisitos:

1. Os Processos Intensivos em Conhecimento são parte integrante de toda organização e, como sua própria denominação, possuem conhecimento fundamental à Gestão do Conhecimento organizacional.
2. Os Processos Intensivos em Conhecimento consistem em ricas fontes de conhecimento tácito, sendo caracterizados pela troca informal de conhecimento entre os participantes através da utilização de ferramentas colaborativas.
3. Os e-mails são fontes ricas para a descoberta de Processos Intensivos em Conhecimento, uma vez que deles podem ser extraídos os elementos necessários para a representação destes processos.
4. A amostra de e-mails a ser selecionada para aplicação do método deve ser composta por e-mails que estejam diretamente relacionados ao Processo Intensivo em Conhecimento que se deseja descobrir.
5. O Processo Intensivo em Conhecimento descoberto, ao final da execução do método proposto, deve servir para apoiar a Gestão do Conhecimento através da explicitação de grande parte do seu conhecimento inerente.

4.2 Fases do Método

Conforme descrito pelos requisitos da proposta, este método tem como principal objetivo descobrir informações relevantes de um Processo Intensivo em Conhecimento, a partir da análise criteriosa do conteúdo da amostra de e-mails selecionada previamente por um ou mais participantes deste processo.

Para isso, é necessária a extração do conteúdo desses e-mails, a estruturação dos dados extraídos, a mineração do conteúdo textual dos e-mails trocadas pelos participantes, a análise dos dados obtidos e, finalmente, a representação destes resultados.

Para atender a este objetivo, o método para descoberta de Processos Intensivos em Conhecimento é dividido em cinco passos: selecionar amostra de e-mails, extrair e estruturar dados relevantes, minerar conteúdo textual dos e-mails, minerar dados relevantes e representar Processo Intensivo em Conhecimento. A ilustração abaixo resume as fases executadas pelo método, com suas respectivas atividades:

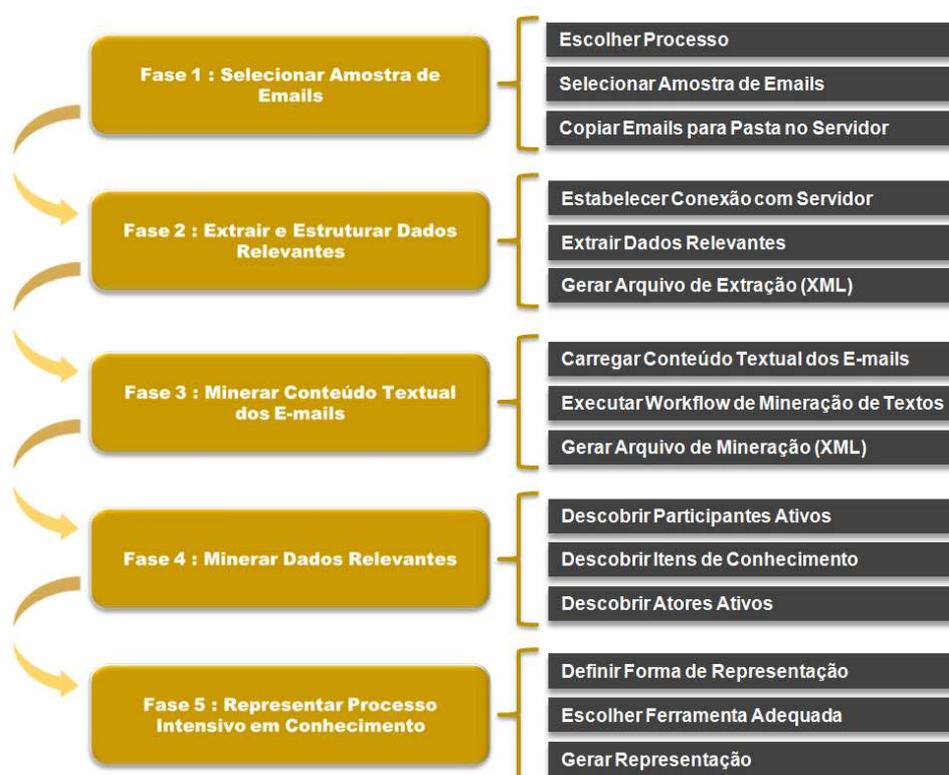


Figura 4.1 – Fases do Método para Descoberta de Processos Intensivos em Conhecimento

4.2.1 Selecionar Amostra de E-mails

A primeira fase do método proposto consiste em uma atividade manual executada por um ou mais participantes do Processo Intensivo em Conhecimento que se deseja descobrir. Esta fase é crucial para que sejam garantidas a qualidade e integridade do resultado a ser obtido futuramente. Para isso, é subdividida em três passos:

- Escolher Processo Intensivo em Conhecimento;
- Selecionar Amostra de e-mails;
- Copiar e-mails para Pasta no Servidor.

4.2.1.1 Escolher Processo Intensivo em Conhecimento

Primeiramente, deve-se escolher o Processo Intensivo em Conhecimento ao qual será aplicado o método proposto. Assume-se que seja um processo relevante ao negócio praticado na Organização em que é executado, de forma que a explicitação do seu conhecimento inerente agregue valor à Gestão do Conhecimento organizacional.

4.2.1.2 Selecionar Amostra de E-mails

Uma vez escolhido o processo a ser analisado, um ou mais participantes deste deverão selecionar um conjunto de e-mails que tenha relacionamento direto com o processo em análise, devendo evitar a escolha de e-mails que tratem de assuntos diversos que possam causar ruído nos resultados extraídos.

4.2.1.3 Copiar E-mails para Pasta no Servidor

Após selecionar a amostra de e-mails, caberá ao participante responsável pela análise, copiá-la para uma pasta específica no servidor de e-mails onde esta possa ser automaticamente localizada. O método proposto oferece integração multiplataforma com quaisquer servidores de e-mails através de protocolos de serviço de mensagens, como SMTP, POP3 e IMAP4.

4.2.2 Extrair e Estruturar Dados Relevantes

Uma vez escolhido o Processo Intensivo em Conhecimento e selecionada a amostra de e-mails a ser analisada, a próxima fase do método propõe a extração e estruturação de todas as informações relevantes contidas nesta amostra, de forma que seja possível aprofundar a análise. Esta fase é executada automaticamente e a solução de seus passos desenvolvida através da tecnologia Java e algumas de suas respectivas APIs (*Application Programming Interfaces*), conforme a necessidade de aplicação. Para fins de automação, esta fase é subdividida em três passos principais:

- Estabelecer conexão com servidor;
- Extrair Dados Relevantes;
- Gerar Arquivo de Extração.

4.2.2.1 Estabelecer Conexão com Servidor

Este primeiro passo trata da conexão com o servidor de e-mails, de forma que seja possível localizar e carregar todas as informações contidas na amostra de e-mails selecionada na fase anterior. Para tal, faz-se necessária a informação dos seguintes dados: endereço do servidor de e-mails, credenciais de acesso ao domínio (*login* e senha) e pasta onde se encontram os e-mails selecionados. Para estabelecer conexão com o servidor, a solução proposta utiliza a *Java Mail API*, uma interface de programação Java independente responsável por prover as classes necessárias para modelar um sistema de e-mails, permitindo assim que a solução possa ser integrada a quaisquer servidores de e-mails.

4.2.2.2 Extrair Dados Relevantes

Neste passo, tal como no passo anterior, é adotada a *Java Mail API*, uma vez que esta oferece diversos recursos para a leitura de mensagens de diversos tipos e formatos, o que torna a solução multiplataforma. Nesta fase, são extraídas todas as informações relevantes a futuras análises, sendo estas: Data de Envio, Remetente, Destinatário(s), Copiado(s), Arquivo(s) Anexo(s), Assunto e Mensagem.

Tais dados servem como base o método proposto, uma vez que permitem a descoberta de elementos fundamentais para a representação de Processos Intensivos em Conhecimento. Para obtenção de melhores resultados, foi desenvolvido um tratamento ao conteúdo textual dos e-mails, através da utilização técnicas de remoção de *Stop Words* (FELDMAN & SANGER, 2007), de forma que textos sem valores semânticos fossem removidos e não gerassem ruídos em análises futuras.

4.2.2.3 Gerar Arquivo de Extração

Uma vez extraídas todas as informações relevantes à análise, é necessário que estas sejam estruturadas, de forma que possam ser lidas e interpretadas pela próxima fase do método. Para atender a esta necessidade, o método proposto adota a utilização do recurso XML para estruturar os dados extraídos.

Para esta solução, nesta fase é utilizada a *JAXB API (Java Architecture for XML Binding)*, para realizar a criação e acesso a documentos XML a partir de aplicações Java. Para tal, foi elaborado um XSD (*XML Schema Definition*), disponível no Apêndice A, para definir a estrutura dos dados extraídos nesta fase do método. A figura 4.2 ilustra a estrutura definida por esse XSD:

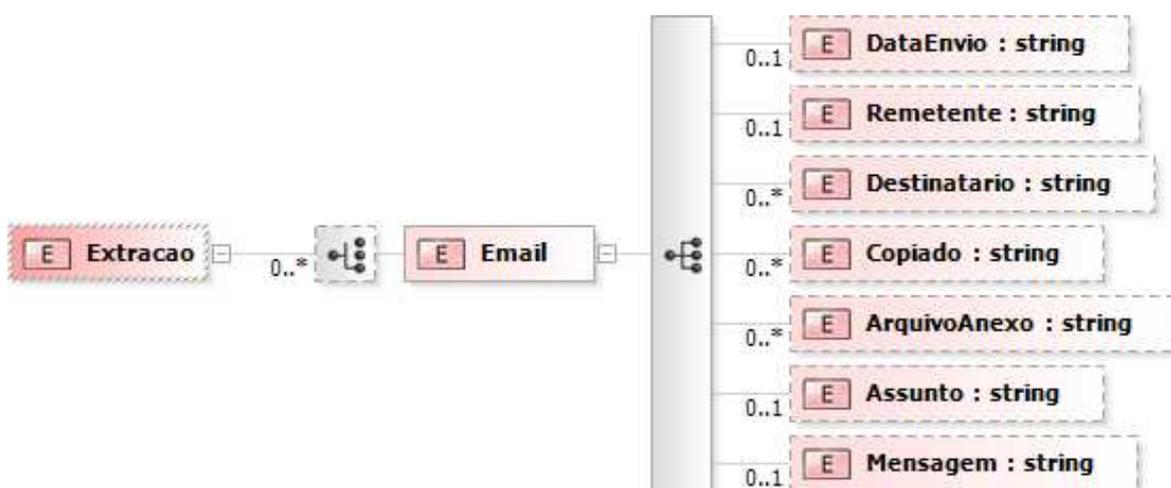


Figura 4.2 – Estrutura do XSD após Extração

4.2.3 Minerar Conteúdo Textual dos E-mails

Uma das informações mais relevantes para a descoberta de Processos Intensivos em Conhecimento é a Atividade, pois é através desta que se torna possível compreender os passos e ações necessários para executar um determinado processo. Apesar de sua relevância, esta não é uma informação disponível simplesmente a partir da extração executada na fase anterior. O resultado gerado previamente traz apenas o conteúdo textual de todos os e-mails contidos na amostra selecionada originalmente.

Contudo, por se tratar de um meio de troca informal de conhecimento, não seria prudente assumir que o texto original contido neste resultado seria suficiente para a descoberta e compreensão do Processo Intensivo em Conhecimento em análise. Por essa razão, é necessário que este conteúdo de texto não estruturado seja analisado, de forma que seja possível extrair, a partir deste, Atividades Candidatas.

Para isso, nesta fase é adotada uma adaptação ao método para eliciação e modelagem de Processos de Negócio *Story Mining*, apresentado no Capítulo Três. Este método apresenta uma solução para a eliciação de Processos de Negócio a partir da mineração de narrativas (textos não estruturados) elaboradas pelos próprios envolvidos nos processos.

Analogamente, a solução proposta por esta pesquisa assume que o conteúdo textual contido e-mails selecionados é semelhante às narrativas aplicadas ao *Story Mining*, uma vez que ambos são textos não estruturados. De mesmo modo, em ambos os casos, deseja-se descobrir as Atividades Candidatas que integram o texto original analisado.

Por essa razão, acredita-se que a aplicação do *Story Mining* ao método proposto viabilizará a descoberta de algumas Atividades existentes no Processo Intensivo em Conhecimento em análise, com base no conteúdo textual dos e-mails originais.

Esta fase utilizará como insumo o arquivo XML gerado pela fase anterior, sendo necessário aplicar a este uma série de passos para fins de automação. Estes passos são:

- Carregar Conteúdo Textual dos E-mails;
- Executar Workflow de Mineração de Textos;
- Gerar Arquivo de Mineração.

4.2.3.1 Carregar Conteúdo Textual dos E-mails

O método *Story Mining* utiliza como insumo textos não estruturados, em formato de narrativas. Por essa razão, este primeiro passo consiste na alteração do Workflow utilizado no desenvolvimento do método *Story Mining* para converter o arquivo XML, conforme estrutura descrita na fase anterior, em classes da linguagem Python. Para atender a este objetivo, neste passo é criado e integrado o componente *XMLParser* ao workflow de mineração de textos, conforme ilustrado na Figura 4.3.

4.2.3.2 Executar Workflow de Mineração de Textos

Neste passo é executado o workflow de mineração de textos que, a partir do arquivo XML carregado no passo anterior, irá minerar as Atividades Candidatas nele contidas. Para tal, devido à grande variedade de técnicas, algoritmos e programas a serem utilizados, o método *Story Mining* adota um Sistema de Gerenciamento de Workflows Científicos (*SGWfC*) para apoiar a definição e execução do workflow de mineração de textos. Um workflow científico pode ser definido como a combinação de dados e sequência de programas (CAVALCANTI, TARGINO, BAIÃO *et al.*, 2005). Um Sistema de Gerenciamento de Workflows Científicos é uma ferramenta que dá apoio computacional à definição, execução e análise comparativa entre as diversas instâncias do workflow, onde podem ser aplicadas técnicas, programas, valores de argumentos de cada algoritmo e fontes de conhecimento para o método como um todo.

Conforme apresentado detalhadamente no Capítulo Três, para a aplicação das técnicas de Mineração de Texto, o método *Story Mining* adota o *SGWfC VisTrails* (CALLAHAN, FREIRE E SANTOS, 2006), devido às suas funcionalidades avançadas para visualização e comparação de resultado, além de sua flexibilidade nativa para integrar funções e aplicações externas, por meio de módulos extras, que podem ser desenvolvidos como classes em linguagem Python e inseridos no *VisTrails* como bibliotecas extras de módulos para workflows científicos.

Para as técnicas de Mineração de Texto e Processamento da Linguagem Natural, o método *Story Mining* utiliza o framework NLTK – *Natural Language Toolkit* (BIRD & LOPER, 2004). Este framework foi adotado devido à vasta gama de algoritmos disponíveis, como *Parsing Sintático*, *Shallow Parsing* e *Tokenização*, além de disponibilizar diversos *corpora* de documentos etiquetados em múltiplos idiomas.

De forma a atender as necessidades desta pesquisa, foram realizadas algumas adaptações ao workflow de minerações de texto proposto pelo método *Story Mining*. No escopo desta pesquisa, foi necessária a, já citada, criação do módulo *XMLParser*, tal como a criação dos módulos *EmailChunkConverter* e *EmailXMLWriter*, os quais serão apresentados a seguir. A Figura 4.3 ilustra o workflow de mineração de textos adaptado a esta pesquisa, conforme exibido na interface do *VisTrails*:

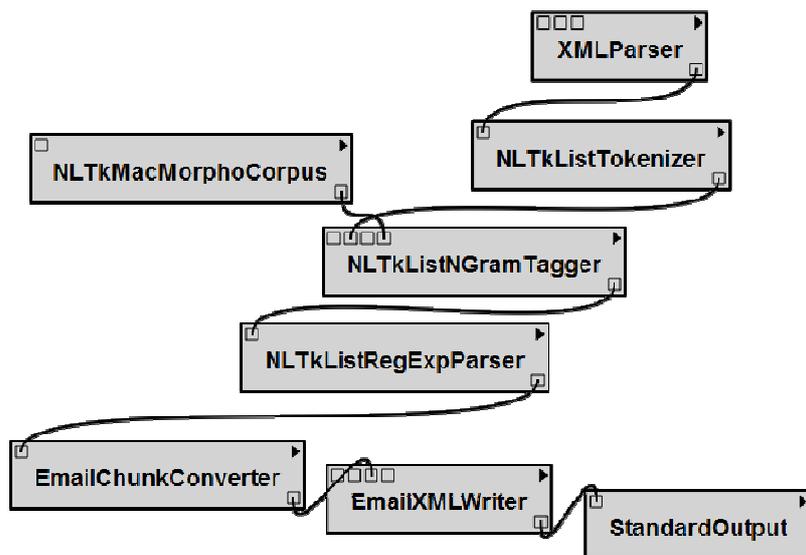


Figura 4.3 – Workflow Científico Adaptado do Método *Story Mining*

Uma vez carregado, no passo anterior, o arquivo XML recebido, este é submetido a um processo típico de processamento da linguagem natural, sendo este iniciado com a Tokenização do conteúdo da mensagem de cada e-mail (*NLTKListTokenizer*), dividindo-o em listas de palavras e sinais característicos.

Depois de concluída a fase de Tokenização, é executada a classificação gramatical automatizada das palavras, através da técnica *Trigram Tagger* (*NLTKListNGramTagger*). Para isso, é utilizado o corpus para a língua portuguesa MAC-MORPHO (ALUISIO, PINHEIRO, MANFRIM, 2004), o qual servirá como dados de treinamento para o algoritmo de classificação.

Após esta fase, os padrões gramaticais são extraídos a partir da técnica de *Shallow Parsing* (OSBORNE, 2000), com base em uma gramática de expressões regulares (*NLTKListRegExpParser*).

4.2.3.3 Gerar Arquivo de Mineração

Uma vez descobertas as Atividades Candidatas, é fundamental que estas sejam estruturadas para que possam ser utilizadas pelo método proposto. Por essa razão, foi necessária uma nova adaptação no método *Story Mining*, uma vez que o resultado gerado pelo workflow de mineração de textos consiste em um arquivo XPDL (*XML Process Definition Language*), utilizado para a representação de Processos de Negócio. O *Story Mining* foi adaptado de modo que as Atividades Candidatas mineradas sejam consolidadas em uma única estrutura de dados (*EmailChunkConverter*), a qual é transformada em um arquivo XML (*EmailXMLWriter*).

Para isso, foi elaborado um novo XSD (XML Schema Definition), disponível no Apêndice A, para definir a estrutura dos dados extraídos juntamente com as Atividades descobertas nesta fase do método. A figura 4.4 ilustra a estrutura definida por esse XSD:

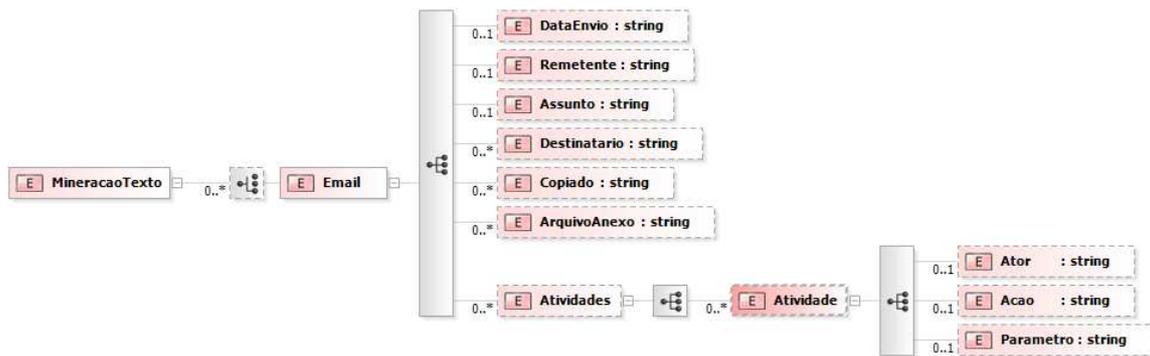


Figura 4.4 – Estrutura do XSD após Mineração de Texto (*Story Mining*)

4.2.4 Minerar Dados Relevantes

Uma vez descobertas as Atividades Candidatas do Processo Intensivo em Conhecimento, é necessária a realização de outras análises que permitam a descoberta dos demais elementos referentes ao processo analisado. Para isso, esta fase tem como objetivo aplicar técnicas de mineração aos dados extraídos, de forma que seja possível descobrir as demais informações relevantes aos Processos Intensivos em Conhecimento, como, por exemplo, os Participantes e suas interações sociais e Itens de Conhecimento.

Esta fase é executada automaticamente e a solução de seus passos desenvolvida através da tecnologia Java, utilizando como massa de dados, o arquivo XML gerado pela fase anterior. Os seus passos estão subdivididos conforme as análises que se deseja realizar, sendo estas:

- Descobrir Participantes Ativos;
- Descobrir Itens de Conhecimento;
- Descobrir Atores Principais;

4.2.4.1 Descobrir Participantes Ativos

Neste passo, são descobertos os Participantes Ativos do processo em análise. São considerados Participantes Ativos todos aqueles que atuam ativamente na execução dos Processos Intensivos em Conhecimento, ou seja, aqueles que enviam e-mails relacionados ao processo analisado.

No contexto desta solução, adotou-se como premissa que para ser ativo basta que o Participante do processo seja o remetente de um dos e-mails selecionados, pois, desta forma, estará atuando ativamente na execução de uma das instâncias deste processo. Uma vez definido o critério para seleção de Participantes Ativos, é necessário analisar qual o peso de atuação de cada Participante, ativo ou não, no processo em questão.

O Peso de Atuação dos participantes de um Processo Intensivo em Conhecimento consiste em um classificador numérico responsável por atribuir uma determinada pontuação aos participantes do processo. Esta pontuação tem como finalidade estabelecer, através da análise comparativa, quais os participantes mais envolvidos no processo. Da mesma forma, este peso de atuação pode ser atribuído ao relacionamento entre participantes, classificando assim o grau de interação social entre os participantes do processo. Para descobrir esta informação, são atribuídos pontos distintos a cada tipo de participação dos envolvidos:

- **Remetente:** a este participante é atribuído o maior peso (3), uma vez que, ao enviar um e-mail relacionado ao processo em questão, apresenta-se como Participante Ativo no contexto deste.
- **Destinatário:** a este participante é atribuído peso dois (2), uma vez que, ao receber diretamente um e-mail relacionado ao processo em questão, apresenta-se como parte do contexto deste.
- **Copiado:** a este participante é atribuído peso um (1), uma vez que, ao ser copiado em um e-mail relacionado ao processo em questão, apresenta-se como possível interessado, mas não necessariamente um participante atuante no contexto deste processo.

Com base na pontuação definida, é realizada uma contagem para definir o peso total de cada participante no processo. Após a contagem de pesos de cada participante, é realizada uma análise para averiguar com quais participantes cada Participante Ativo se relaciona, sendo executada novamente uma contagem de pesos que irá medir o grau de relacionamento de cada Participante Ativo com os demais participantes do processo.

Ao final deste passo, é possível extrair um mapa da rede social existente no processo em análise, capaz de informar os Participantes Ativos do processo e o peso de participação de cada um, os participantes com quem se relacionam e o grau de interação existente entre eles.

4.2.4.2 Descobrir Itens de Conhecimento

Neste passo, são descobertos os Itens de Conhecimento do processo em análise. São considerados Itens de Conhecimento todos os termos que estejam diretamente relacionados ao conhecimento inerente ao Processo Intensivo em Conhecimento. No contexto desta solução assume-se que Itens de Conhecimento são os termos mais frequentes no conteúdo textual e assuntos dos e-mails.

Para isso, é realizada uma contagem entre todo o conteúdo textual dos e-mails selecionados, de forma a identificar quais termos mais se repetem no escopo de atuação do processo em questão. Nesta lógica foi aplicado o mesmo tratamento de palavras de parada (*Stop Words*) utilizado anteriormente na fase dois, para evitar a seleção de termos da língua portuguesa irrelevantes para o contexto que se deseja descobrir, como pronomes, artigos, preposições, etc. Uma vez descobertos, os Itens de Conhecimento serão associados ao processo em análise.

4.2.4.3 Descobrir Atores Principais

Neste passo, são descobertos os Atores Principais de cada Atividade Candidata descoberta na fase anterior. É importante frisar que cada atividade poderá conter, caso identificado, um ator. São considerados Atores Principais todos os atores das Atividades Candidatas que fizerem algum tipo de referência a um ou mais participantes do processo.

Para isso, é realizada uma análise em cada Atividade Candidata descoberta, de forma a identificar se o ator existente possui relação direta com o nome específico de algum participante, de forma completa ou parcial. Da mesma forma, é realizada uma análise que visa averiguar se o ator identificado faz menção ao grupo de participantes, assumindo assim que todos os participantes daquela interação serão Atores Principais.

Vale ressaltar que são considerados Atores Ativos um ou mais participantes de um processo que estejam relacionados à execução de uma Atividade Candidata, enquanto os Participantes Ativos são todos os participantes do processo que enviam e-mails relacionados ao processo em questão.

4.2.5 Representar Processo Intensivo em Conhecimento

Uma vez selecionada uma amostra de e-mails referente a um Processo Intensivo em Conhecimento, extraídos os dados originais relevantes, mineradas as Atividades Candidatas que integram todo o conteúdo textual dos e-mails trocados e executadas as análises que auxiliam e complementam a interpretação deste processo, é necessário definir uma abordagem para a sua representação. Esta fase é executada automaticamente e a solução de seus passos desenvolvida através da tecnologia Java, utilizando como entrada, os dados extraídos e analisados pelas fases anteriores. Para fins de automação, esta fase é subdividida em três passos:

- Definir Forma de Representação;
- Escolher Ferramenta Adequada;
- Gerar Representação.

4.2.5.1 Definir Forma de Representação

Conforme apresentado no Capítulo Dois, apesar de haver inúmeras propostas na literatura, não há uma forma oficial ou mais adotada para a representação de Processos Intensivos em Conhecimento. Sendo assim, a solução proposta por este método irá adotar um recurso existente, conhecido e flexível, que permita representar, de forma clara, objetiva e intuitiva, os elementos que compõem o processo analisado.

Vale ressaltar que esta proposta de solução objetiva apoiar a Gestão do Conhecimento através da descoberta de novos conhecimentos inerentes a Processos Intensivos em Conhecimento existentes. Portanto, visando a atender às necessidades da solução proposta e oferecer uma forma de representação familiar à área de Gestão do Conhecimento, acredita-se que a utilização de Mapas Mentais (BUZAN, 1996) seja uma estratégia interessante a ser adotada, uma vez que atende às principais necessidades apresentadas por este método.

Segundo BUZAN (1996), Mapa mental (*Mindmap*) é o nome dado para um tipo de diagrama, voltado para a Gestão de Informações, Conhecimento e Capital Intelectual. Pode ser aplicado na compreensão e solução de problemas, como ferramenta de brainstorming, no auxílio da gestão estratégica de uma empresa ou negócio, entre outros.

Os desenhos feitos em um mapa mental partem de um ponto central, a partir do qual são irradiadas as informações relacionadas. Podem ser elaborados por meio de canetas coloridas sobre folhas de papel ou uma ferramenta computacional dedicada. Pode ser aplicado a qualquer tarefa, atividade, profissional, ou lazer, de modo individual ou em grupo para planejar qualquer tipo de evento.

Trata-se de um método para planejamento e registro gráfico cada vez mais usado em todas as áreas de conhecimento humano (BUZAN, 1996). Por essa razão, a solução proposta irá adotar Mapas Mentais como forma de representação dos elementos de um Processo Intensivo em Conhecimento.

4.2.5.2 Escolher Ferramenta Adequada

Neste passo é realizada uma pesquisa visando escolher a ferramenta adequada a ser adotada por este método. Uma vez definido o Mapa Mental como forma de representação, foram levantadas as três ferramentas mais utilizadas para a elaboração de Mapas Mentais, sendo estas:

- **MindMan Personal:** ferramenta gratuita e em inglês. Não insere figuras e não oferece recursos importantes.
- **FreeMind:** ferramenta gratuita e em português. Desenvolvida em Java e compatível com vários sistemas operacionais. Permite a utilização de ícones, hyperlinks, envoltórias coloridas em ramos, exportação para página Web com controle de contração e expansão de tópicos e imagens em tópicos.
- **EasyMapper:** ferramenta comercial e em português. Permite a utilização de tópicos mediante digitação simples, formatação de linhas e bordas, imagens de arquivos, hyperlinks em tópicos e exportação para páginas Web.

No escopo de atuação da solução proposta para este método, é adotada a ferramenta *FreeMind* para elaboração dos Mapas Mentais, uma vez que, conforme descrito acima, é gratuita, desenvolvida em Java, o que facilita a integração com este método, em português e oferece todos os recursos necessários para representação completa dos elementos de um Processo Intensivo em Conhecimento.

4.2.5.3 Gerar Representação

Finalmente, este passo é responsável por gerar a representação do processo em análise. Para isso, é necessário que, primeiramente, seja definida uma estrutura básica de Mapa Mental, a partir da qual os processos possam ser diagramados.

Conforme apresentado no Capítulo Dois, para que um diagrama represente um Processo Intensivo em Conhecimento, é necessário que contemple os seguintes elementos: Processo, Participantes e suas Interações Sociais, Atividades, Itens de Conhecimento e Artefatos de Conhecimento.

Estes elementos deverão ser dispostos organizadamente no diagrama visando simplificar a sua navegabilidade e leitura. Por se tratar de um Mapa Mental, a estrutura a ser definida deverá partir de um único ponto central, que no contexto desta solução será o Processo Intensivo em Conhecimento. A partir deste ponto, deverão ser exibidos todos os Itens de Conhecimento citados, os Artefatos trocados e os Participantes Ativos, cada qual com seu respectivo peso no processo, segundo analisado na fase anterior. Uma vez expandido, cada Participante Ativo possuirá duas perspectivas:

- Participantes com quem interage;
- Atividades que executa.

Caso a navegação ocorra por Participantes com quem interage, deverão ser listados todos os Participantes com quem o Participante Ativo expandido se relaciona, exibindo, para cada qual, o grau de interação entre ambos, ou seja, o peso analisado na fase anterior. Para cada Participante, deverão ser listadas as datas em que ocorreram as interações, agrupadas por mês/ano e dia/hora de forma que seja possível oferecer uma visão temporal das interações ocorridas entre os participantes. Para cada interação, deverão ser listadas as Atividades executadas, os Itens de Conhecimento citados e os Artefatos trocados.

Caso a navegação ocorra por Atividades que executa, deverão ser listadas as datas em que ocorreram as interações, agrupadas por mês/ano e dia/hora de forma que seja possível oferecer uma visão temporal das interações ocorridas. Para cada interação, deverão ser listadas as Atividades executadas, os Itens de Conhecimento citados, os Artefatos trocados e, principalmente, os Participantes com quem o Participante Ativo expandido se relacionou nesta específica interação.

Vale ressaltar que para cada Atividade executada, poderão ser listados Itens de Conhecimento citados e Atores Principais, caso existam. Desta forma, a estrutura definida permite que a navegabilidade seja realizada pelo Mapa Mental de distintas formas, oferecendo múltiplas perspectivas do Processo Intensivo em Conhecimento, permitindo uma visão panorâmica e completa do processo que se deseja descobrir. A figura 4.5 ilustra uma visão básica expandida da estrutura definida:

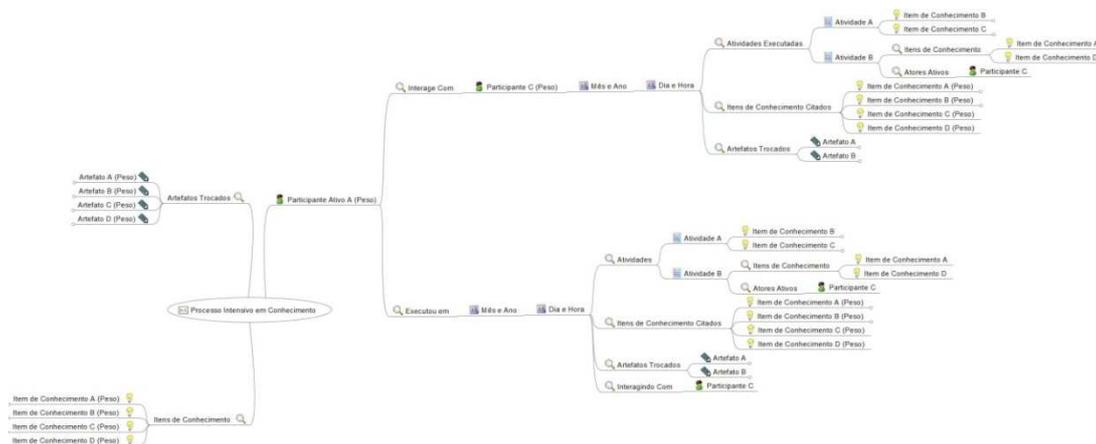


Figura 4.5 – Visão Básica Expandida da Estrutura do Mapa Mental

Uma vez definida a estrutura do Mapa Mental que irá representar os elementos do processo em questão, resta elaborar uma estratégia para gerar automaticamente estes diagramas. Ao criar Mapas Mentais através da ferramenta *FreeMind*, são gerados arquivos de extensão proprietária *.mm (*Mind Map*). Por se tratar de uma ferramenta desenvolvida em Java, gratuita e de código-fonte aberto, os arquivos por ela gerados, nada mais são, que arquivos XML cujo esquema se encontra disponível e acessível.

Ao analisar o XSD encontrado, disponível no Apêndice A, é possível perceber que o esquema adotado pelo *FreeMind* baseia-se na estrutura de dados de Grafos, onde cada elemento exibido consiste em um novo nó, o qual sempre possuirá um nó pai e poderá ou não possuir nós filhos. Cada nó irá conter atributos específicos que irão caracterizar o elemento, como: texto, fonte, ícone, posição no diagrama, entre outros. A figura 4.6 ilustra a estrutura do XSD utilizado pelo *FreeMind*:

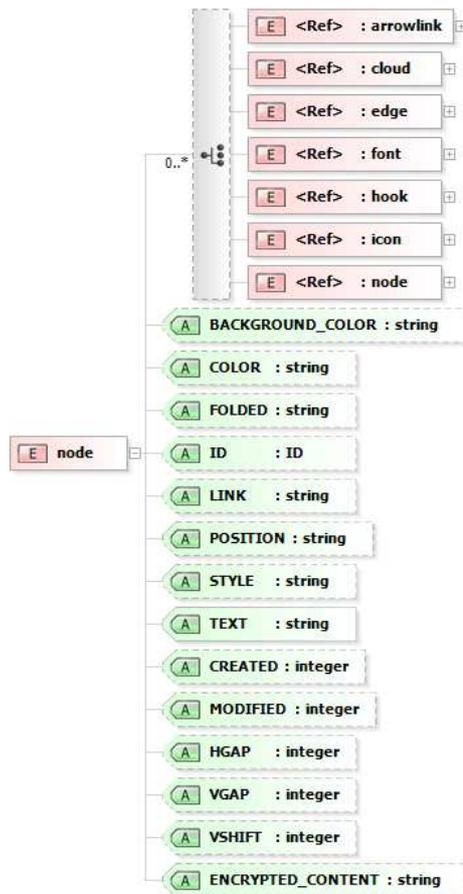


Figura 4.6 – Estrutura do XSD Utilizado pelo *FreeMind*

Depois de analisada a estrutura de XSD adotada pelo *FreeMind*, resta executar o passo que irá, finalmente, representar o processo em questão. Este passo deverá ser executado automaticamente e a solução desenvolvida através da tecnologia Java. Para isso, assim como nas fases anteriores, foi utilizada a *JAXB API (Java Architecture for XML Binding)*, para a criação de classes Java com base no XSD apresentado na Figura 4.6. A partir destas classes, é possível criar os devidos nós, segundo a estrutura de Mapa Mental definida neste passo, e atribuir a estes os valores, pesos e características desejados para a representação do Processo Intensivo em Conhecimento.

O resultado final consiste na geração de um arquivo XML, com extensão *.mm, cujo formato poderá ser interpretado pelo *FreeMind*, de forma que este renderize e represente o Processo Intensivo em Conhecimento analisado.

Uma vez aberto o arquivo pela ferramenta, esta oferecerá diversos recursos de exportação do diagrama gerado, permitindo a exibição do processo em qualquer computador, tornando a navegação do processo independente de ferramenta. A figura 4.7 ilustra a navegação por um Processo Intensivo em Conhecimento através de um navegador Web:

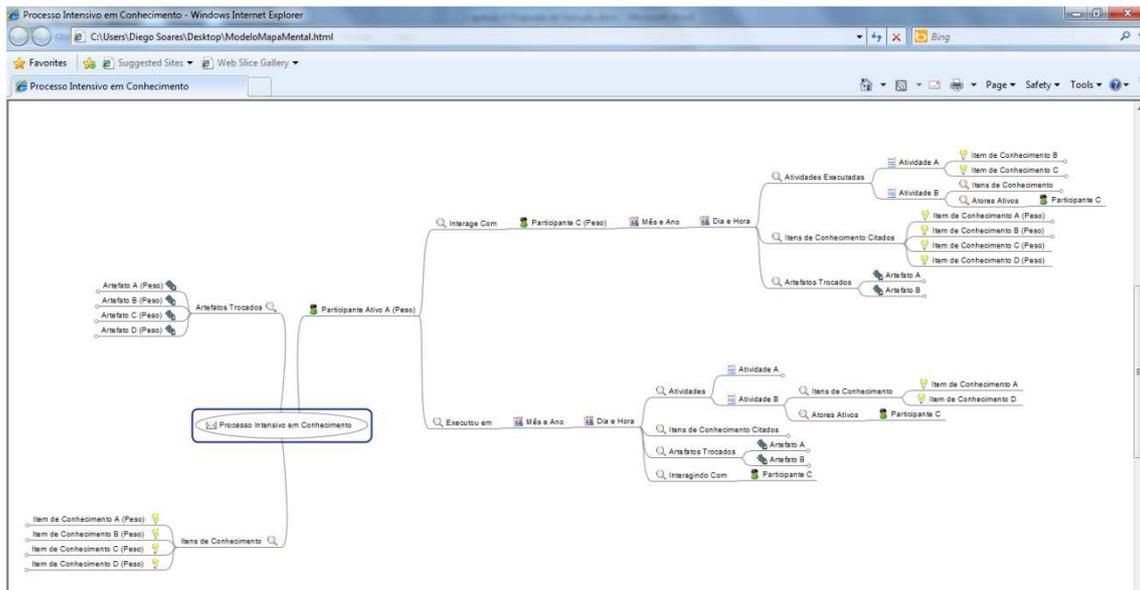


Figura 4.7 – Processo Intensivo em Conhecimento em um Navegador Web

Uma vez concluída a construção da solução proposta para o método de descoberta de Processos Intensivos em Conhecimento, este necessita ser executado em um contexto de mundo real, de forma que possa ser devidamente avaliado, através do método de pesquisa do Estudo de Caso, cujo planejamento e resultados obtidos serão apresentados a seguir.

Capítulo 5 – Estudo de Caso e Análise de Resultados

Este capítulo descreve o planejamento adotado para realização do Estudo de Caso e análise dos resultados obtidos. Esta aplicação tem como objetivo avaliar o método de descoberta de Processos Intensivos em Conhecimento proposto em um contexto real de uma empresa de Mídia e Entretenimento para aparelhos móveis.

5.1 Metodologia

O pesquisador Robert K. Yin define o Estudo de Caso como uma análise empírica que investiga um fenômeno contemporâneo dentro de seu contexto de vida real, especialmente quando os limites entre o fenômeno e o contexto não estão claramente definidos (YIN, 1984, p. 23). Em outras palavras, a utilização de Estudos de Caso como método de pesquisa auxilia a compreensão de questões complexas, enfatizando a análise contextual detalhada de um número limitado de eventos ou condições e suas relações.

Por esse motivo, há muitos anos, pesquisadores de diversas disciplinas têm adotado este método de pesquisa. Em particular, as áreas de Ciência da Computação e Sistemas de Informação têm feito amplo uso desse método de pesquisa qualitativa para examinar situações contemporâneas da vida real, visando fornecer a base para a aplicação de suas ideias e métodos de extensão.

Nesta pesquisa será adotado o Estudo de Caso como método de pesquisa para avaliar as propriedades do método proposto e as condições nas quais se comporta. Portanto, para avaliar a validade e confiabilidade da solução proposta, foram planejados dois tipos distintos de estudo: Estudo Exploratório e Estudo de Caso.

O Estudo Exploratório teve como principal objetivo realizar a exploração do método proposto, o que permitiu observar detalhadamente o seu comportamento durante a execução das suas diversas fases e, a partir destas, avaliar os resultados obtidos. Para isso, foi utilizado um cenário real ao qual foi aplicado o método proposto. Por questões de confidencialidade das informações envolvidas, não foi possível a divulgação completa dos resultados obtidos neste estudo. Estas análises prévias auxiliaram o levantamento de questões importantes para a pesquisa e ressaltaram a necessidade de tratamentos especiais que permitiram a melhor visualização e navegação pelo Processo Intensivo em Conhecimento descoberto. Ao todo, para a realização do Estudo Exploratório foram necessárias duas semanas. Para a execução do método, foram necessárias 3 horas e 20 minutos. Para a análise dos resultados obtidos, foram necessárias 2 horas e 45 minutos. Para a correção de falhas e melhorias da solução, foi necessária, aproximadamente, uma semana, de forma que o método estivesse preparado para ser aplicado ao Estudo de Caso.

O Estudo de Caso, conforme descrito anteriormente, teve como objetivo avaliar a validade do método proposto por esta pesquisa, aplicando-o a um cenário real de uma organização. Uma vez concluída a execução da solução, os resultados obtidos foram analisados pelos principais participantes do Processo Intensivo em Conhecimento selecionado, através da realização de uma entrevista. Ao todo, para a realização do Estudo de Caso foi necessário, aproximadamente, um mês. A seleção da amostra de e-mails a ser utilizada foi realizada pelos próprios participantes do processo e, para isso, foram necessários 10 dias. Para a execução do método, foi necessária 1 hora e 15 minutos, não sendo necessárias alterações, adaptações e/ou correções na solução. Para a preparação e realização da entrevista para análise dos resultados obtidos, foram necessárias 3 horas e 50 minutos.

Nas seções seguintes, serão apresentados detalhadamente os cenários, os passos da execução do método proposto e os resultados obtidos a partir dos estudos realizados.

5.2 Estudo Exploratório – Processo “Importar Mídias”

Para averiguar se o método proposto por esta pesquisa atende aos requisitos listados na proposta de solução, foi necessária, primeiramente, a realização de um Estudo Exploratório. Este estudo teve como principal objetivo analisar o resultado gerado pelo método proposto a partir de um Processo Intensivo em Conhecimento de escopo conhecido, portanto selecionado e analisado pelo próprio pesquisador, o que possibilitou uma avaliação abrangente do resultado gerado.

5.2.1 Descrição do Cenário

Em uma empresa de Mídia e Entretenimento para aparelhos móveis, há um projeto de TI cujo principal foco trata da venda *online* de mídias, sendo estas: imagens, vídeos, músicas, *ringtones*, entre outros. Uma das maiores metas deste projeto trata do aumento exponencial do acervo de mídias, o que tornaria este serviço o maior do Brasil.

Para disponibilização destas mídias, esta empresa dispõe de inúmeros contratos com diversas gravadoras, sendo estas responsáveis pelo envio de milhares de mídias mensalmente, visando o aumento significativo do acervo de mídias disponível para vendas. Para tal, cabe à supracitada empresa prover um processo responsável por:

- ✓ Recebimento das mídias enviadas;
- ✓ Conversão dos formatos de arquivos das mídias;
- ✓ Aplicação de proteção para Gestão de Direitos Digitais;
- ✓ Persistência dos metadados em Banco de Dados;
- ✓ Publicação para venda.

Este processo é denominado Importar Mídias, sendo executado diariamente e responsável por publicar milhares de novas mídias por dia. Conseqüentemente, assume-se que este processo agrega enorme valor para a organização, uma vez que, através da execução deste, é possível alcançar uma das metas principais da empresa.

Por essa razão, é fundamental que o conhecimento inerente a este processo integre a Gestão do Conhecimento desta organização. No entanto, apesar das atividades executadas serem conhecidas em um nível macro, não são conhecidas as tarefas necessárias para a execução deste processo.

A sequência com que estas tarefas são executadas alterna a cada instância do processo, conforme tomada de decisões dos envolvidos responsáveis, onde conhecimento, experiência e criatividade tornam-se peças fundamentais para o seu correto funcionamento. Logo, para ser executado, o processo Importar Mídias depende diretamente da tomada de decisão dos seus envolvidos, fato este que o caracteriza como um Processo Intensivo em Conhecimento. Por esse motivo, foi adotado este processo para o Estudo Exploratório desta pesquisa.

5.2.2 Execução

Primeiramente, foi realizada a atividade manual “Selecionar Amostra de E-mails” do método proposto, onde, para aplicação de um teste de carga, foi selecionada, pelo pesquisador, uma amostra significativa contendo 730 e-mails relacionados ao Processo Intensivo em Conhecimento Importar Mídias.

Estes e-mails foram selecionados com base na análise do conteúdo textual, tal como nos assuntos relacionados, de forma que fosse detectada sua relação direta com o processo que se deseja descobrir.

Após selecionada, a amostra de e-mails foi devidamente movida para uma pasta no servidor de e-mails denominada Importar Mídias. Dado o elevado volume de e-mails selecionados, foram necessárias duas horas para concluir esta atividade.

Em seguida, foi executada a fase “Extrair Dados Relevantes” do método proposto, com base no endereço do servidor de e-mails, credenciais de acesso (login e senha) e local no servidor onde se encontravam os e-mails selecionados.

Foram necessários 6 minutos para que o método estabelecesse conexão com o servidor, carregasse todos os e-mails contidos na amostra, extraísse informações relevantes para análises futuras, conforme definido no capítulo anterior, e gerasse o arquivo XML de extração, composto por 7570 linhas e tenha como tamanho 680 KB. A figura 5.1 ilustra parcialmente o resultado gerado por esta fase:

```

<Email>
  <DataEnvio>15 Janeiro 2010</DataEnvio>
  <Remetente>Thiago Guedes Cardoso Tudela</Remetente>
  <Destinatario>Bernardo Filardi</Destinatario>
  <Destinatario>Fabio Eduardo Toneto Nunes</Destinatario>
  <Copiado>Diego Soares</Copiado>
  <Copiado>Andre Galvani Oliveira</Copiado>
  <Assunto>Máquina 172.16.0.78 está fora do ar (IMPORTAÇÃO DE MÍDIAS)</Assunto>
  <Mensagem>Bernardo, Conforme falamos, por gentileza, veja se existe algum bloqueio de firewall em sua rede interna. Vou pedir também, a gentileza de efetuar um teste numa conexão direta (como 3G ou adsl). A janela que deverá abrir é semelhante a esta:</Mensagem>
</Email>

```

Figura 5.1 – Exemplo de Arquivo XML Gerado pela Fase de Extração

Concluída a geração do arquivo XML da fase de extração, torna-se possível a execução da próxima fase do método proposto, “Minerar Conteúdo Textual dos E-mails”. Para isso, foi executado pelo *VisTrails* o workflow científico de mineração de textos, após informado o local onde o arquivo XML de extração foi gerado. Foi necessária 1 hora e 15 minutos para execução das técnicas de Tokenização, classificação gramatical das palavras, extração de padrões gramaticais baseada em uma gramática de expressões regulares, consolidação das atividades descobertas e geração do arquivo XML de mineração. A figura 5.2 ilustra parcialmente o resultado gerado por esta fase:

```

<Email>
  <DataEnvio>15 Janeiro 2010</DataEnvio>
  <Remetente>Thiago Guedes Cardoso Tudela</Remetente>
  <Destinatario>Bernardo Filardi</Destinatario>
  <Destinatario>Fabio Eduardo Toneto Nunes</Destinatario>
  <Copiado>Diego Soares</Copiado>
  <Copiado>Andre Galvani Oliveira</Copiado>
  <Assunto>Máquina 172.16.0.78 está fora do ar (IMPORTAÇÃO DE MÍDIAS)</Assunto>
  <Atividades>
    <Atividade>
      <Ator>Conforme</Ator>
      <Acao>falamos</Acao>
      <Parametro>, por gentileza , veja se existe algum bloqueio de firewall em sua rede interna . </Parametro>
    </Atividade>
    <Atividade>
      <Ator/>
      <Acao>Vou pedir</Acao>
      <Parametro>também, a gentileza de efetuar um teste numa conexão direta ( </Parametro>
    </Atividade>
  </Atividades>
</Email>

```

Figura 5.2 – Exemplo de Arquivo XML Gerado pela Fase de Mineração de Texto

O arquivo XML gerado serviu como entrada para a próxima fase do método proposto, “Minerar Dados Relevantes”. Esta fase tem como principal objetivo extrair informações relevantes que complementem e auxiliem a melhor compreensão do Processo Intensivo em Conhecimento que se deseja descobrir.

Conforme apresentado no Capítulo anterior, durante esta fase foram descobertos os Participantes Ativos do processo, o peso de participação de cada um, os Participantes com quem cada Participante Ativo se relacionou, o grau de interação entre os Participantes, os Itens de Conhecimento citados ao longo das interações e os Artefatos trocados. Para estas análises, foram necessários, aproximadamente, 4 minutos de processamento e os resultados obtidos mantidos em memória para a execução da próxima fase.

A próxima e última fase do método proposto teve como principal objetivo “Representar Processo Intensivo em Conhecimento”. Para isso, esta fase foi executada logo após o processamento da fase anterior, de forma que os resultados disponíveis pudessem ser aplicados diretamente às classes Java geradas para representar o Mapa Mental. Foram necessários 2 minutos e 45 segundos de processamento para instanciar os elementos, aplicar os valores analisados e gerar o arquivo XML. A figura 5.3 ilustra parcialmente o resultado gerado por esta fase:

```
<node TEXT=">Vou pedir também, a gentileza de efetuar um teste numa conexão direta." FOLDED="true">
  <icon BUILTIN="list"/>
  <node TEXT="Atores Principais" FOLDED="true">
    <icon BUILTIN="xmag"/>
    <node TEXT="Thiago Guedes Cardoso Tudela(2)" FOLDED="true">
      <icon BUILTIN="male2"/>
    </node>
    <node TEXT="Fabio Eduardo Toneto Nunes(2)" FOLDED="true">
      <icon BUILTIN="male2"/>
    </node>
    <node TEXT="Diego Soares(1)" FOLDED="true">
      <icon BUILTIN="male2"/>
    </node>
    <node TEXT="Andre Galvani Oliveira(1)" FOLDED="true">
      <icon BUILTIN="male2"/>
    </node>
  </node>
  <node TEXT="Itens de Conhecimento Citados" FOLDED="true">
    <icon BUILTIN="xmag"/>
    <node TEXT="resposta" FOLDED="true">
      <icon BUILTIN="idea"/>
    </node>
  </node>
</node>
```

Figura 5.3 – Exemplo de Arquivo XML Gerado pela Fase de Representação

O arquivo XML gerado foi estruturado de forma que fosse capaz de ser lido e interpretado pela ferramenta de edição de Mapas Mentais escolhida, *FreeMind*. A Figura 5.4 ilustra o resultado final obtido, uma vez renderizado pela ferramenta:

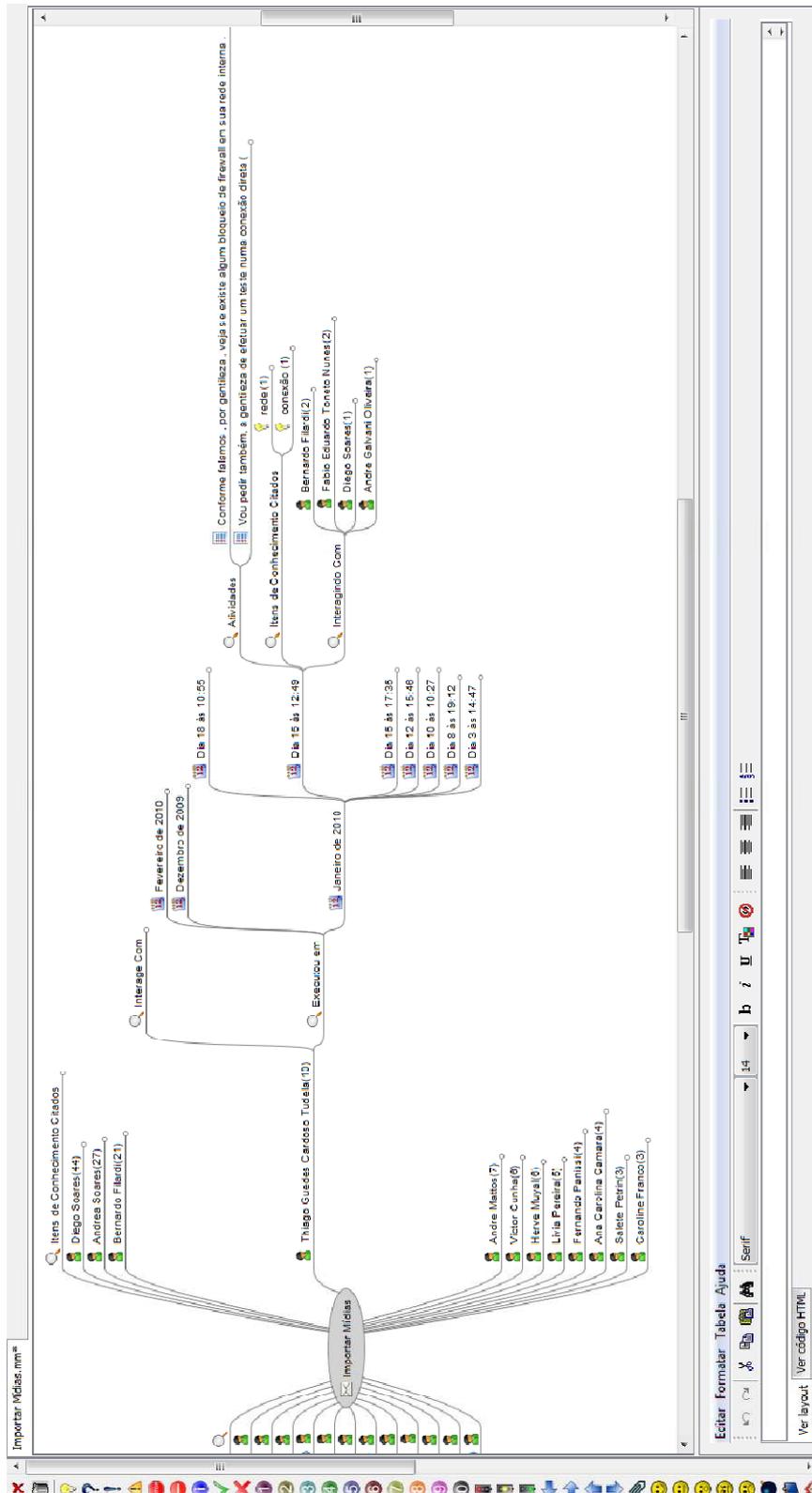


Figura 5.4 – Representação do Processo Importar Mídias

5.2.3 Análise dos Resultados

Uma vez concluída a execução do método proposto, o resultado obtido foi cautelosamente analisado pelo pesquisador, uma vez que este conhece o contexto no qual o processo é executado, já que foi um Participante Ativo deste. Esta análise permitiu avaliar parcialmente se o método se encontrava apto a ser aplicado ao Estudo de Caso.

5.2.3.1 Participantes Ativos

Em primeiro lugar, foi analisada a relação de Participantes Ativos descoberta, visando identificar se esta reflete a realidade do processo em questão. O processo Importar Mídias é executado pelos seguintes papéis: Representante da Gravadora, Assistente de Marketing, Gerente de Projetos de TI, Analista de Sistemas e Analista de Infraestrutura. Todos estes papéis são atuantes no processo, portanto devem ter seus responsáveis listados na relação de Participantes Ativos.

Foi constatado que todos os responsáveis pelos papéis existentes no processo foram devidamente listados na relação de Participantes Ativos, sendo, inclusive, os pesos sugeridos coerentes com as respectivas participações. Ao analisar os pesos sugeridos, foi possível perceber que os Analistas de Sistemas e Analistas de Infraestrutura possuem maior peso de participação no processo. Esta descoberta confirma que estes participantes são aqueles realmente envolvidos na maioria das atividades deste processo.

Da mesma forma, ficou claro que os Assistentes de Marketing, tal como os Gerentes de Projetos de TI, possuem participação média no processo. Esta descoberta também reflete a realidade, uma vez que estes participantes se envolvem no processo apenas para garantir sua plena execução, de forma que suas atividades no processo tratam apenas do seu acompanhamento. Os Representantes das Gravadoras ficaram com menor peso, uma vez que suas atividades no processo tratam apenas de eventuais problemas ou dúvidas sobre o processo.

5.2.3.2 Relacionamento entre Participantes

Foi realizada uma análise de forma a verificar se o relacionamento entre os Participantes do Processo refletiu a interação social praticada na realidade. Ao analisar o Mapa Mental é possível descobrir que os Analistas de Sistemas têm maior interação com Gerente de Projetos de TI e os Analistas de Infraestrutura. Isto caracteriza uma verdade, uma vez que seus trabalhos são diretamente dependentes e interligados.

Também foi possível perceber, a partir da representação, que os Analistas de Marketing têm maior interação com os Analistas de Sistemas e Representantes da Gravadora. Isso reflete a realidade, uma vez que os Analistas de Marketing realizam a interface entre as demandas das gravadoras e os responsáveis pelo funcionamento do processo. Dessa forma, ficou evidente que o nível de interação sugerido entre os Participantes corresponde ao ocorrido na prática durante a execução do Processo.

5.2.3.3 Atividades Candidatas

Ao analisar as Atividades Candidatas sugeridas como executadas pelos Participantes, foi possível constatar que para a maioria das atividades foram descobertos corretamente os seus Atores Principais. Esta informação permitiu que fossem descobertas rapidamente os responsáveis pela execução das atividades, sem que houvesse a necessidade de interpretações mais aprofundadas.

Além disso, quanto às atividades executadas pelos Analistas de Sistemas, na maioria dos casos, pode-se perceber que estas tratam de correções de falhas, desenvolvimento de melhorias, investigações de problemas de Infraestrutura e publicação de mídias urgentes. Da mesma forma, fica evidente que os Analistas de Infraestrutura executam atividades relacionadas à investigação de falhas de acesso aos servidores, liberação de permissão de acesso às redes, realização de manutenções preventivas e acompanhamento de chamados. Estas descobertas caracterizam um fato, uma vez que os Analistas de Infraestrutura e os Analistas de Sistemas são os responsáveis técnicos pelo pleno funcionamento do processo.

Também foi possível visualizar, a partir da análise do resultado obtido, que tanto os Gerentes de Projetos de TI quanto os Analistas de Marketing executam, na maioria dos casos, atividades relacionadas ao acompanhamento de problemas, solicitação de publicação de lançamentos, atendimento a demandas e elaboração de cronogramas e prazos. Estas descobertas refletem a realidade, uma vez que a participação dos Gerentes de Projetos de TI e dos Analistas de Marketing trata do controle da execução do processo, sendo suas atividades relacionadas diretamente ao acompanhamento deste. Logo, as Atividades Candidatas sugeridas refletem as atividades executadas no cotidiano por todos os papéis envolvidos no processo analisado.

5.2.3.4 Artefatos Gerados

A partir da análise da relação de Artefatos Gerados ao longo das interações, e seus respectivos pesos, foi possível identificar que grande parte destes Artefatos é gerada pelos Analistas de Infraestrutura. Estes artefatos têm como objetivo documentar informações como configuração de servidores e arquitetura de Infraestrutura, o que ressalta o fato de que há muito conhecimento espalhado em torno das atividades realizadas. Esta descoberta reflete uma verdade, já que é uma prática dos Analistas de Infraestrutura manter documentadas as informações importantes referentes à configuração da Infraestrutura, de forma que sirvam como base de instrução para outros grupos.

Nesta mesma análise, foi identificado que os Analistas de Marketing são responsáveis pela geração de inúmeros Artefatos ao longo do processo, com base nos diversos documentos de planos de campanha referentes a novos lançamentos de mídias descobertos pelo método proposto. Este fato também demonstra uma nova realidade do processo, uma vez que uma das principais responsabilidades dos Analistas de Marketing trata da criação de novas ideias e campanhas para acelerar a venda de mídias, o que envolve a geração de documentos, que, normalmente,

acabam perdidos pela organização. Com base nessa interpretação do resultado, foi possível assumir que os artefatos descobertos são relevantes ao processo e, portanto, podem ser utilizados como fonte de conhecimento sobre este.

Dada à relevância das informações contidas nos Artefatos descobertos, uma possível proposta futura seria o armazenamento automático dos artefatos gerados pelos Participantes Ativos do processo. Ou seja, uma vez descobertos pelo método proposto, estes Artefatos poderiam ser armazenados em um Repositório de Conhecimento Organizacional, evitando-se assim que o conhecimento inerente a estes artefatos ficasse espalhado pela organização e garantindo a centralização desta informação competitiva.

5.2.3.5 Itens de Conhecimento

Finalmente, foram analisados os 55 Itens de Conhecimento descobertos pelo método proposto. Nesta análise, foram considerados os itens de conhecimento mais citados, ou seja, aqueles cujas contagens de frequência resultaram em uma pontuação maior, uma vez que estes podem ser interpretados como descobertas de maior relevância. Foi possível denotar que termos comuns ao processo, como Mídia, Importação, Música, Artista, Álbum, UPC, ISRC, Gravadoras, entre outros, foram os mais citados em todas as interações analisadas. Segundo a análise realizada, os Itens de Conhecimento contidos neste resultado refletiram informações importantes para a contextualização do processo, pois são termos do domínio utilizados pelos participantes.

Também foi possível notar que Itens de Conhecimento referentes à Infraestrutura foram muito citados, como Servidor, FTP, Banco, Dados, IP, Conexão, Acesso, entre outros, o que ressalta, uma vez mais, a forte dependência que o processo apresenta em relação à Infraestrutura. Por último, foi identificado que todas as gravadoras tornaram-se Itens de Conhecimento, o que ressalta a importância das Gravadoras no processo Importar Mídias.

5.2.4 Conclusão

Este Estudo Exploratório, através da análise apresentada na seção anterior, permite uma avaliação parcial satisfatória do método proposto, dado que este atende, nesta primeira análise, aos requisitos listados na proposta de solução. Através deste estudo, foi possível denotar que há conhecimento relevante inerente ao Processo Intensivo em Conhecimento analisado na amostra de e-mails selecionada e que, a partir destes, é possível descobrir os elementos necessários para a representação destes processos.

Através da leitura do Mapa Mental do processo Importar Mídias, gerado pelo método proposto, foi possível descobrir e aprender informações relevantes à execução deste processo, que anteriormente se encontravam apenas dispostas no conhecimento tácito dos participantes envolvidos.

Vale ressaltar que esta conclusão reflete o ponto de vista de apenas um Participante Ativo deste processo. Possivelmente, caso esta análise tivesse sido realizada por outro Participante Ativo, novas leituras pudessem ter sido realizadas, tal como outras informações não descobertas.

Por essa razão, para uma avaliação completa e abrangente do método proposto, sob todos os pontos de vista, faz-se necessário o envolvimento de dos principais participantes do processo, desde a seleção da amostra de e-mails a análise dos resultados obtidos. Por essa razão, foi realizada uma segunda análise imparcial, a ser detalhadamente descrita a seguir.

5.3 Estudo de Caso – Processo “Monitorar Envio de Notícias”

Uma vez executado o Estudo Exploratório e, conseqüentemente, avaliada a estabilidade da solução proposta, foi realizado o Estudo de Caso para avaliar o método proposto por esta pesquisa, aplicando-o a um cenário real de uma organização. Dessa forma, foi possível avaliar se a proposta de solução atende ou não à hipótese apresentada.

5.3.1 Descrição

Em uma empresa de Mídia e Entretenimento para aparelhos móveis, há um projeto de TI cujo principal foco trata de um serviço de envio de notícias via SMS (*Short Message Service* ou Serviço de Mensagens Curtas). Este serviço oferece a seus assinantes o recurso de receber, diariamente, diversas notícias relacionadas aos canais assinados.

Este serviço dispõe de, aproximadamente, três milhões de assinantes ativos, ou seja, assinantes que pagam para receber as mensagens enviadas. Ao longo de um dia, são cadastradas centenas de novas notícias, sendo estas atribuídas aos seus respectivos canais que, ao todo, somam cerca de 110 canais disponíveis para assinatura. Cada canal cadastrado possui seus respectivos horários de envio, sendo estes horários distribuídos ao longo do dia de forma a evitar concorrência de envio de notícias em determinados períodos.

Ao todo, são enviadas, pelo menos, nove milhões de notícias diariamente, o que o torna um dos mais importantes serviços de VAS (*Value-added Services* ou Serviços de Valor Agregado) da organização, representando parte significativa da receita bruta dos serviços oferecidos e crescente a cada ano. Devido à sua importância na organização, é fundamental que haja um controle efetivo deste serviço, de forma a garantir que as notícias não deixem de ser enviadas, que ocorram atrasos nos envios ou que sejam enviadas duplicadamente. Para atender a este objetivo, foi criado o processo Monitorar Envio de Notícias.

Este processo consiste no acompanhamento diário dos alarmes de envio de notícias. Estes alarmes são consultas, aplicadas ao Banco de Dados, orquestradas periodicamente por um sistema de monitoramento visando à identificação de inconsistências. Uma vez encontrado um sinal de falha no envio de notícias, este sistema irá soar um alarme personalizado e encaminhar a mensagem de erro automaticamente, via e-mails, a todos os participantes envolvidos no processo.

Este é considerado o evento de início do processo Monitorar Envio de Notícias, o qual pode ser dividido nas seguintes atividades:

- ✓ Acompanhar alarmes gerados pelo Sistema de Monitoramento;
- ✓ Encaminhar erros para análise;
- ✓ Investigar possível causa do problema;
- ✓ Corrigir, se necessário, falha identificada;
- ✓ Reportar solução aplicada ao erro encontrado.

Este processo é considerado crucial para a empresa, uma vez que, através dele, é possível garantir a saúde e integridade de um dos principais serviços de valor agregados por ela oferecidos. Por essa razão, é fundamental que o conhecimento inerente a este processo integre a Gestão do Conhecimento desta organização. No entanto, apesar das atividades executadas serem conhecidas em um nível macro, não são conhecidas as tarefas necessárias para a execução deste processo.

A sequência de análises realizada para descobrir a causa de uma falha no envio de uma notícia, por exemplo, é alternada a cada instância do processo, conforme tomada de decisões dos envolvidos responsáveis, onde conhecimento, experiência e criatividade são peças fundamentais para o seu correto funcionamento. Portanto, para ser executado, o processo Monitorar Envio de Notícias depende diretamente da tomada de decisão dos seus envolvidos, fato este que o caracteriza como um Processo Intensivo em Conhecimento. Por esse motivo, foi adotado este processo para Estudo de Caso nesta pesquisa.

5.3.2 Execução

Primeiramente, foi realizada a atividade manual “Selecionar Amostra de E-mails” do método proposto, onde os principais participantes do processo selecionaram uma amostra contendo 97 e-mails relacionados, com base na análise do conteúdo textual, tal como nos assuntos relacionados, de forma que fosse detectada sua relação direta com o processo que se deseja descobrir. Após selecionados, estes e-mails foram devidamente movidos para uma pasta no servidor de e-mails denominada Monitorar Envio de Notícias. Devido à ausência de tempo disponível dos participantes, foram necessários 10 dias para que esta atividade fosse concluída.

Em seguida, foi executada a fase “Extrair Dados Relevantes” do método proposto, com base no endereço do servidor de e-mails, credenciais de acesso de um dos participantes (login e senha) e local no servidor onde se encontravam os e-mails selecionados. Foram necessários 2 minutos e 45 segundos para estabelecer conexão com o servidor, carregar todos os e-mails contidos na amostra, extrair as informações relevantes para análises futuras, conforme definido no capítulo anterior, e gerar o arquivo XML de extração.

Concluída a geração do arquivo XML da fase de extração, torna-se possível a execução da próxima fase do método proposto, “Minerar Conteúdo Textual dos E-mails”. Para isso, foi executado, pelo *VisTrails*, o workflow científico de mineração de textos, após informado o local onde o arquivo XML de extração foi gerado. Foram necessários 40 minutos para a execução das técnicas de Tokenização, classificação gramatical das palavras, extração de padrões gramaticais baseada em uma gramática de expressões regulares, consolidação das atividades descobertas e geração do arquivo XML de mineração.

O arquivo XML gerado serviu como entrada para a próxima fase do método proposto, “Minerar Dados Relevantes”. Esta fase teve como principal objetivo extrair informações relevantes que complementassem e auxiliassem a melhor compreensão do Processo Intensivo em Conhecimento que se desejava descobrir. Conforme apresentado no Capítulo anterior, durante esta fase devem ser descobertos os Participantes Ativos do processo, o peso de participação de cada um, os Participantes com quem cada Participante Ativo se relacionou, o grau de interação entre os Participantes, os Itens de Conhecimento citados ao longo das interações e os Artefatos trocados. Para estas análises, foram necessários 2 minutos e 15 segundos de processamento e os resultados obtidos mantidos em memória para a execução da próxima fase.

A próxima e última fase do método proposto teve como principal objetivo “Representar Processo Intensivo em Conhecimento”. Para isso, esta fase foi executada logo após o processamento da fase anterior, de forma que os resultados disponíveis pudessem ser aplicados diretamente às classes Java geradas para representar o Mapa Mental. Foram necessários 55 segundos de processamento para instanciar os elementos, aplicar os valores analisados e gerar o arquivo XML, disponível no Apêndice B. O arquivo XML gerado foi estruturado de forma que fosse capaz de ser lido e interpretado pela ferramenta de edição de Mapas Mentais escolhida, *FreeMind*.

5.3.3 Análise dos Resultados

Para a análise dos resultados obtidos neste Estudo de Caso, foi realizada uma entrevista com os principais participantes do processo Monitorar Envio de Notícias, de forma que os elementos descobertos referentes ao Processo Intensivo em Conhecimento pudessem ser avaliados. A preparação da entrevista teve duração de uma hora e para a sua realização foram necessárias 2 horas e 45 minutos.

De forma que a entrevista fosse conduzida visando à avaliação detalhada dos elementos do processo analisado, foi elaborado um *script* que permitiu a ordenação das perguntas a serem realizadas, tal como incentivou comentários pertinentes por parte dos participantes. A figura 5.5 exibe o *script* aplicado à entrevista:

<p>- Abertura (30 minutos)</p> <p>Agradecer a participação de todos e pedir-lhes que descrevam o Processo Monitorar Envio de Notícias, as atividades executadas e os papéis envolvidos.</p> <p>- Apresentar o Mapa Mental (20 minutos)</p> <p>Apresentar detalhadamente o resultado obtido a partir do método proposto, dando destaque aos elementos do Processo Intensivo em Conhecimento: Participantes Ativos, Relacionamentos entre Participantes, Artefatos Gerados e Itens de Conhecimento Citados.</p> <p>- Realizar Perguntas (60 minutos)</p> <ol style="list-style-type: none">1. A relação de Participantes Ativos descoberta pelo método reflete a realidade do Processo em questão?2. Os pesos sugeridos para cada Participante Ativo são coerentes com as respectivas participações?3. O relacionamento entre os Participantes do Processo reflete a interação social praticada na realidade?4. O nível de interação sugerido entre os Participantes corresponde ao ocorrido no dia-a-dia do Processo?5. As Atividades sugeridas como executadas pelos Participantes procedem, ou seja, representam ações ou passos executados no processo em análise?6. Os Atores Principais sugeridos às Atividades são coerentes?7. Os Artefatos Gerados listados são coerentes?8. A relação de Itens de Conhecimento listada é coerente à realidade, ou seja, representa o tipo de conhecimento envolvido no processo? <p>- Comentários Finais (20 minutos)</p> <p>Nesta fase final, pedir-lhes para que descrevam suas opiniões, críticas e/ ou elogios com relação ao resultado apresentado, como: facilidade de interpretação, compreensão, estrutura e organização da representação do Processo.</p>

Figura 5.5 – Script da Entrevista de Estudo de Caso

Estiveram presentes na entrevista três participantes do processo Monitorar Envio de Notícias, os quais apresentaram os papéis existentes e suas respectivas responsabilidades:

- ✓ **Analista de Infraestrutura:** responsável por acompanhar o sistema de monitoramento de envio de notícias. Ao soar um alarme, este participante analisa a mensagem de erro. Caso identifique que não há notícia cadastrada, cabe a este participante encaminhar o erro ao Jornalista. Caso não identifique a causa do erro, cabe a este participante encaminhar a mensagem de erro ao Analista de Marketing.
- ✓ **Analista de Marketing:** responsável por garantir que todas as notícias sejam enviadas pontualmente, sem atrasos ou falhas. Cabe a este participante acompanhar os erros ocorridos durante o envio de notícias e suas respectivas soluções junto ao Analista de Sistemas.
- ✓ **Analista de Sistemas:** responsável por realizar a análise criteriosa do erro. Caso o erro seja causado por falhas no sistema de envio ou na consulta realizada, cabe a este participante realizar a sua correção. Caso o erro seja causado por ausência de notícias cadastradas, cabe a este participante notificar ao Jornalista encarregado. Uma vez solucionado o problema, cabe a este participante reportar a solução dada à ocorrência ao Analista de Marketing.
- ✓ **Jornalista:** responsável por cadastrar diariamente as centenas de notícias associadas aos canais disponíveis para assinatura.

Estavam presentes na entrevista os três representantes dos papéis mais atuantes no processo: Analista de Infraestrutura, Analista de Marketing e Analista de Sistemas. Uma vez tendo estes se apresentado, descrito detalhadamente o processo, os papéis existentes e as respectivas responsabilidades, foi-lhes apresentado o Mapa Mental gerado pelo método proposto com base na amostra de e-mails por eles selecionada. Com base nesta apresentação, foram iniciadas as devidas análises.

5.3.3.1 Participantes Ativos

A relação de Participantes Ativos descoberta pelo método proposto foi analisada pelos participantes do processo, que identificaram que todos os Participantes Ativos descobertos de fato executam diretamente o processo analisado.

O Analista de Infraestrutura atentou ao fato do grupo “Operação” ser classificado como Participante Ativo, uma vez que este consiste em um grupo de pessoas onde todas possuem o mesmo papel no processo, o de Analistas de Infraestrutura. Justificou ainda que, como o monitoramento de envio de notícias ocorre 24x7, todos os analistas responsáveis pela Infraestrutura devem acompanhar os incidentes ocorridos. Por esta razão, o grupo é envolvido na maioria das atividades do processo e, conseqüentemente, para o método proposto, foi classificado como Participante Ativo do processo.

A Analista de Marketing atentou ao fato de os estagiários das respectivas áreas terem sido classificados como Participantes Ativos, uma vez que não são oficialmente Analistas. No entanto, para o método proposto, por participarem ativamente do processo, estes são atuam como Participantes Ativos. Nesta mesma análise, o Analista de Sistemas ressaltou que, apesar de serem classificados como Participantes Ativos, os estagiários foram pontuados, pelo método proposto, com um menor peso de participação. Para ele, esta pontuação reflete a realidade, uma vez que não são todos os incidentes que são atendidos pelos estagiários.

Ainda com relação à pontuação dos pesos, o Analista de Sistemas ressaltou que a sugestão do método é coerente com a prática realizada. Para justificar sua assertiva, explicou que os incidentes são comunicados pelo Analista de Infraestrutura e, muitas vezes, pelo Analista de Marketing, mas sempre são atendidos pelo Analista de Sistemas. Ao somar o peso de participação do Analista de Infraestrutura (9) com o peso de participação da Analista de Marketing (10) e comparar o resultado com o peso de participação do Analista de Sistemas (18), concluiu que a sugestão de pesos é coerente e reflete a realidade praticada.

5.3.3.2 Relacionamento entre Participantes

O relacionamento entre os participantes foi mais simples de analisar, visto que, durante a apresentação dos papéis e responsabilidades do processo, foram descritas as interações entre os papéis, conforme resumido na tabela abaixo:

	Analista de Infraestrutura	Analista de Marketing	Analista de Sistemas	Jornalista
Analista de Infraestrutura		X		X
Analista de Marketing			X	
Analista de Sistemas	X	X		X
Jornalista		X	X	

Tabela 5.1 – Relacionamento entre Papéis no Processo Monitorar Envio de Notícias

Com base nesta informação, provida ao início da entrevista, foi necessária apenas a validação no Mapa Mental apresentado. O Analista de Infraestrutura identificou que sempre se relaciona com os grupos Operação, Jornalismo e Interatividade. Explicou que o grupo Interatividade pertence à equipe de Marketing e concluiu afirmando que o relacionamento entre participantes sugerido é verdadeiro.

Ao analisar os seus relacionamentos, a Analista de Marketing notou que, exceto as interações realizadas com a outra Analista de Marketing, todos os participantes com quem se relaciona são Analistas de Sistemas. Inclusive, ressaltou que o participante com maior pontuação de peso de interação é o Analista de Sistemas principal responsável pelo processo, o que reflete a sua realidade.

O Analista de Sistemas percebeu que se relaciona com todos os participantes do processo, ressaltando que, por ser o responsável por resolver todos os incidentes, na prática diária, se relaciona com participantes de todos os papéis do processo. Ainda complementou que, apesar de interagir com todos os participantes do processo, seus relacionamentos de maior peso são com Analistas de Marketing, o que é justificado pelo fato de ter que reportar todas as soluções a estes participantes.

5.3.3.3 Atividades Candidatas

Ao analisar as Atividades Candidatas sugeridas pelo método proposto, o Analista de Infraestrutura percebeu que, em todas as atividades por ele executadas, sempre há a ação de comunicar a ocorrência de um erro no envio de notícias, como: “*Estamos **como** alarme abaixo, poderiam verificar?*”. Explicou que, na maioria dos casos, sua responsabilidade se resume a comunicar às Analistas de Marketing sobre a ocorrência de erros no envio de notícias. Conclui que, de fato, estes são os passos realizados por ele no processo.

A Analista de Marketing, ao analisar as atividades sugeridas, identificou que, em muitos casos, as atividades por ela executadas tratam de garantir se o envio de notícias está ocorrendo nos horários corretos, como: “*as notícias estão postadas no PubNews com o horário certinho. Estamos com algum problema?*” ou “*O horário dele não pode ser 13h , já que deveria estar sintonizado com os horários do canal Coritiba que são 15h30 e 19h30*”. Da mesma forma, verificou que, em diversas situações, acompanha a solução de problemas, como em: “*peço! Isso foi consertado?*”. Lembrou sua responsabilidade no processo, uma vez que deve garantir que o envio de notícias ocorra normalmente. Por essa razão, ao final de sua avaliação, concluiu que as atividades sugeridas refletem a sua participação no processo.

O Analista de Sistema iniciou sua análise ressaltando que, através da maioria das atividades, fica evidente sua participação nas soluções dos incidentes, como em: “*Tem algum problema nesse monitoramento que **nós ainda** estamos investigando*”. Ressaltou também que muitos dos problemas ocorrem devido a falhas de acesso a banco, como descrito na atividade: “*Esse alarme corresponde ao problema de acesso ao banco que tivemos hoje às 9:30*”. Informou ainda que, para detecção de novos erros, é o responsável por elaborar novas consultas ao Banco de Dados, como fica evidente na atividade: “*estou criando um monitoramento novo para termos alarmes **quandoos** envios começarem a atrasar*”.

O Analista de Sistemas concluiu sua análise destacando um erro no método proposto, onde em diversos casos, conforme destacados acima, algumas palavras são exibidas juntas, como nos exemplos: *como* ao invés de *com o*, *nósainda* ao invés de *nós ainda e quandoos* ao invés de *quando os*. Este problema foi investigado e detectada a falha durante a execução da fase de Tokenização do workflow de mineração de textos.

Além disso, o Analista de Sistemas sugeriu uma proposta de trabalho futuro onde fosse possível, tal como feito com os Participantes, relacionar atividades executadas, de forma a descobrir que, sempre que uma determinada Atividade A é executada, uma outra atividade B é executada a seguir. Desta forma, defendeu o aumento da eficiência na execução diária do processo, uma vez que não seria necessário investigar erros conhecidos.

5.3.3.4 Artefatos Gerados

Ao analisar os Artefatos Gerados pelo processo, o Analista de Infraestrutura identificou que um dos artefatos mais criados ao longo das execuções do processo em questão consiste em uma planilha contendo todas as notícias cujo envio não pode ser realizado. Ressaltou que é sua responsabilidade criar estas planilhas de forma que sejam documentadas todas as notícias não enviadas para futuras análises.

O Analista de Sistemas ratificou que, na maioria dos casos em que ocorre uma falha no envio de notícias, o Analista de Infraestrutura gera a relação de notícias não enviadas. Da mesma forma, identificou que muitos Artefatos Gerados no processo tratam de arquivos contendo scripts SQL. Justificou informando que, sempre que detectada a falha na consulta executada pelo sistema de monitoramento, cabe ao Analista de Sistemas realizar os acertos nos arquivos SQL de consulta e enviá-los para o Analista de Infraestrutura. Sugeriu apenas que os Artefatos Gerados descobertos pelo método proposto fossem ordenados por peso, de forma a facilitar a leitura. Concluiu informando que estes são coerentes com a realidade do processo.

5.3.3.5 Itens de Conhecimento

Ao analisar os Itens de Conhecimento descobertos, o Analista de Infraestrutura destacou alguns termos ligados diretamente ao conhecimento envolvido na execução deste processo, como: *alarme, envio, notícia, horários, monitoramento, canais*, entre outros. Ressaltou o fato de “*banco*” ser classificado como Item de Conhecimento, justificando-o pelo volume elevado de ocorrências de falhas de envio de notícia devido a problemas de acesso ao Banco de Dados.

A Analista de Marketing destacou a constante classificação de times de futebol como Itens de Conhecimento, como: *jogo, esportes, flamengo, botafogo, goiás*, entre outros. Explicou que os times de futebol são os canais de notícias mais assinados e, por essa razão, o monitoramento de falhas no envio de notícias destes canais é mais prioritário.

O Analista de Sistemas complementou lembrando que, para times de futebol, as notícias devem ser enviadas a cada gol e, por essa razão, não são tolerados atrasos nos envios destas notícias. Logo, em sua opinião, todos os Itens de Conhecimento sugeridos representam o escopo de atuação do processo.

5.3.4 Conclusão

Uma vez concluídas as análises, os participantes expressaram suas opiniões a respeito do resultado apresentado. Excetuando as palavras sem separação identificadas e as listas de Artefato e Itens de Conhecimento sem ordenação por peso, todos os participantes concordaram que o resultado obtido reflete a realidade do processo por eles executado. Externaram satisfação com relação à navegação proposta para o processo, tal como à forma rápida e intuitiva oferecida para descobrir e aprender um novo processo. Dessa forma, foi possível concluir que o método elaborado por esta pesquisa atendeu à hipótese proposta, uma vez que, sua validade e confiabilidade puderam ser avaliadas com sucesso por este Estudo de Caso.

Capítulo 6 – Conclusões e Trabalhos Futuros

Intensas transformações sociais, econômicas e tecnológicas vêm redesenhando a atividade produtiva, ao colocar o conhecimento como principal recurso na criação de vantagens competitivas. O conhecimento é uma forma de desenvolvimento cultural e depende de fatores externos ao indivíduo, podendo ser adquirido e compartilhado de diversas maneiras. Portanto, a globalização força a inovação e, conseqüentemente, aponta para a necessidade de Gestão do Conhecimento. A Gestão do Conhecimento consiste em uma área de pesquisa voltada à elaboração de novas estratégias e métodos que auxiliem a criação, aquisição, gerenciamento e, principalmente, materialização do conhecimento na forma de inovação em produtos e serviços.

Um viés da Gestão do Conhecimento considera que o Conhecimento organizacional, geralmente, encontra-se espalhado, não só em documentos ou repositórios, mas também nas rotinas organizacionais, processos, práticas e normas. Sob esta perspectiva, entende-se que a execução de Processos de Negócio constitui uma importante forma de apoio à descoberta de conhecimento. Por essa razão, foi criada uma subárea de pesquisa denominada Gestão do Conhecimento Orientada a Processos de Negócio, responsável por prover metodologias, métodos e ferramentas necessários para atender a este objetivo. Algumas propostas na literatura, como: *BPO-KM*, *PROMOTE*, *CommonKADS*, entre outras, caracterizam tentativas existentes de prover apoio computacional à descoberta de conhecimento a partir de Processos de Negócio.

Neste contexto, é possível denotar a relevância dos Processos Intensivos em Conhecimento, também conhecidos como Processos de Negócio Intensivos em Conhecimento, dada a intensidade de conhecimento presente na execução destes processos por parte de seus participantes. A forte intensidade de conhecimento dos processos de negócio é caracterizada pela dependência direta da criatividade, inovação, experiência e conhecimento dos participantes envolvidos, sendo a execução destes processos baseada em constantes tomadas de decisão. Essas principais características dos Processos de Negócio Intensivos em Conhecimento os fazem ser executados de maneira *ad hoc*. Por esse motivo, o principal objetivo, em longo prazo, da Gestão do Conhecimento Orientada a Processos consiste em identificar, modelar, analisar e aperfeiçoar os Processos Intensivos em Conhecimento. Visando alcançar a este objetivo, inúmeras propostas na literatura, como *BPK*, *DECOR*, *KMDL*, entre outras, sugerem, entre outras coisas, meios para apoiar a descoberta e representação de Processos Intensivos em Conhecimento.

No entanto, mesmo com essas iniciativas, ainda não há linguagens ou ferramentas capazes de separar informação tácita de explícita na representação destes Processos de Negócio. Sendo assim, estas abordagens ainda não permitem visualizar, de maneira adequada, as decisões, ações e métricas envolvidas na sequência de atividades executadas destes processos. Além disso, a intensa socialização e a constante troca informal de conhecimento impedem que o apoio da TI aos Processos Intensivos em Conhecimento seja evoluído adequadamente. Por essa mesma razão, acredita-se que há conhecimento tácito intrínseco no conteúdo textual gerado, através de Ferramentas Colaborativas, pela troca informal de conhecimento.

Esta pesquisa tem como objetivo desenvolver um método para a descoberta automática de Processos Intensivos em Conhecimento com base em uma de suas principais características, a intensa socialização que envolve a execução de suas atividades.

Este método é responsável por extrair, através da mineração de textos, os elementos necessários para a representação de Processos Intensivos em Conhecimento, a partir da troca informal de conhecimento presente no conteúdo textual gerado por uma ferramenta colaborativa, os e-mails. No escopo desta pesquisa, os elementos a serem descobertos são: Processos, Participantes, Interações Sociais, Atividades, Itens de Conhecimento e Artefatos de Conhecimento. Para atender a este objetivo, o método proposto divide-se em cinco fases.

A primeira fase consiste em Selecionar Amostra de E-mails, onde é escolhido o Processo Intensivo em Conhecimento que se deseja descobrir e a amostra de e-mails a este relacionada. A seleção da amostra de e-mails deve envolver os principais Participantes do processo, os quais selecionam, manualmente, alguns e-mails relacionados ao processo a ser analisado.

A segunda fase consiste em Extrair e Estruturar Dados Relevantes, onde a amostra de e-mails selecionada é automaticamente processada, através de módulos desenvolvidos na linguagem Java, de forma que sejam extraídas informações relevantes a análises futuras, gerando ao final um arquivo estruturado XML, de forma que possa ser processado pelas fases vindouras.

A terceira fase consiste em Minerar Conteúdo Textual dos E-mails, onde o conteúdo textual de todos os e-mails selecionados é analisado, de forma que seja possível extrair as Atividades Candidatas que compõem o processo em análise. Para isso, foi adaptado o método *Story Mining*, que trata da elicitación de Processos de Negócio a partir da mineração de narrativas colaborativas (*Group Storytelling*). Para adaptação do método *Story Mining* às necessidades desta pesquisa foi necessária a criação de três novos módulos, elaborados na linguagem *Python*: *XMLParser*, *EmailChunkConverter* e *EmailXMLWriter*, que têm como finalidade, respectivamente: realizar a leitura do arquivo XML de extração, consolidar e estruturar os resultados gerados no processo de mineração de textos e gerar um novo arquivo XML contendo as informações descobertas.

A quarta fase consiste em Minerar Dados Relevantes, onde as informações extraídas originalmente na segunda fase, juntamente com as Atividades descobertas pela adaptação do método *Story Mining*, são automaticamente processadas, através de módulos desenvolvidos na linguagem Java, de forma que os demais elementos do Processo Intensivo em Conhecimento sejam devidamente descobertos. Esta fase considera como Participantes Ativos todos os remetentes dos e-mails selecionados, realizando uma contagem de frequência que permita definir o grau de participação destes no processo. Da mesma forma, são descobertos os demais Participantes do processo e suas interações com os demais Participantes Ativos, realizando uma contagem de frequência que permita ilustrar o grau de interação social entre os participantes. Esta fase considera Artefatos de Conhecimento todos os arquivos anexos trocados durante a execução das atividades do processo, realizando uma contagem de frequência que permita ilustrar a relevância dos artefatos durante a execução do processo analisado. Esta fase considera Itens de Conhecimento os termos mais citados durante as interações realizadas na execução do processo, realizando uma contagem de frequência que permita ilustrar os itens de domínio que caracterizam o processo em questão.

A quinta e última fase do processo consiste em Representar Processos Intensivos em Conhecimento, onde todos os elementos descobertos nas fases anteriores são automaticamente estruturados, através de módulos desenvolvidos na linguagem Java, de forma que possam ser visualizados graficamente. Para isso, foi realizado um estudo na literatura de propostas para a representação de Processos Intensivos em Conhecimento. No entanto, ainda não há uma proposta adequada à adoção por esta pesquisa. Por essa razão, a solução proposta adotou um recurso gráfico conhecido e flexível, o Mapa Mental, um método para planejamento e registro gráfico cada vez mais usado em todas as áreas de conhecimento humano. Para a representação automatizada do Mapa Mental, foi adotada a ferramenta *FreeMind*, devido a robustez e facilidade de integração oferecidos.

Para avaliação do método proposto, foram realizados dois estudos distintos: um Estudo Exploratório e um Estudo de Caso. O Estudo Exploratório teve como objetivo realizar uma avaliação prévia operacional do método proposto, de forma a garantir que as fases elaboradas são capazes de gerar o resultado esperado. Para isso, o próprio pesquisador selecionou, a partir de um cenário organizacional real, uma amostra composta por 730 e-mails, relacionada a um Processo Intensivo de Conhecimento de escopo por ele conhecido, denominado Importar Mídias. Esta análise, apesar de parcial, permitiu a análise detalhada dos resultados obtidos, de forma que eventuais falhas ou melhorias pudessem ser realizadas. Além disso, através desta, foi possível denotar se, na visão do pesquisador, o método cumpre com o esperado.

Para uma aplicação mais abrangente e imparcial do método proposto, foi realizado um Estudo de Caso envolvendo os três principais participantes de um Processo Intensivo em Conhecimento organizacional denominado Monitorar Envio de Notícias. Para isso, o grupo de participantes selecionou uma amostra composta por 97 e-mails relacionada ao processo a ser analisado. O método proposto foi aplicado a esta amostra de e-mails e o resultado obtido pode ser avaliado durante uma entrevista com a participação dos principais envolvidos no processo. Apesar de algumas críticas, como a falha ao separar algumas palavras durante a mineração de textos e a ausência de ordenação ao listar os Artefatos e Itens de Conhecimento, a análise realizada por todos os participantes permitiu uma avaliação satisfatória do método proposto, uma vez que todos os elementos descobertos refletem a realidade dos participantes durante a execução do Processo Intensivo em Conhecimento.

Dessa forma, pode-se concluir que o método proposto cumpre com a hipótese original de descobrir, a partir do conteúdo textual contido em ferramentas colaborativas, os elementos necessários para a representação de Processos Intensivos em Conhecimento.

6.1 Contribuições Realizadas

A principal contribuição desta pesquisa consiste no método para a descoberta de Processos Intensivos em Conhecimento a partir de uma ferramenta colaborativa, os e-mails. Isto caracteriza uma contribuição uma vez que, apesar de existirem na literatura, propostas para a descoberta de Processos Intensivos em Conhecimento, conforme citadas nesta dissertação, não há uma proposta que permita a descoberta automática destes processos, tal como a representação gráfica dos elementos descobertos. Dessa forma, o método proposto apresenta-se como um apoio significativo ao avanço do apoio da TI à Gestão do Conhecimento. Além disso, outras contribuições resultantes do trabalho são:

- O estudo e definição de elementos relevantes à representação de Processos Intensivos em Conhecimento;
- A adaptação do método *Story Mining* através do desenvolvimento de novos módulos específicos para o tratamento de arquivos XML;
- A aplicação do conceito de Mapas Mentais à representação dos elementos relevantes aos Processos Intensivos em Conhecimento;
- Os resultados dos estudos de casos, apresentando evidências sobre a forte intensidade de conhecimento existente durante a execução de Processos Intensivos em Conhecimento, a partir da análise dos Participantes envolvidos.

6.2 Limitações da Pesquisa

Uma limitação da presente pesquisa consiste na seleção manual da amostra de e-mails a ser analisada, uma vez que, apesar de realizada envolvendo os principais participantes do processo, alguns dos e-mails selecionados poderão conter informações referentes a questões fora do contexto do processo analisado, os quais irão, eventualmente, causar ruídos, como, por exemplo, a descoberta de Atividades Candidatas que não pertencem ao processo analisado.

Outra limitação desta pesquisa consiste na não utilização do conceito dos *Threads* relacionados aos emails. Uma possível aplicação a ser realizada com este conceito seria a utilização de *Threads* como agrupadores dos emails analisados, de forma a identificar automaticamente, por exemplo, a sequência de atividades executadas no processo. Na presente pesquisa, esta informação é descoberta a partir da análise temporal das atividades executadas.

6.3 Trabalhos Futuros

O método proposto por esta pesquisa apresenta-se como uma solução para a descoberta de conhecimento inerente a execução de Processos de Negócio caracterizados como intensivos em conhecimento, dada a forte intensidade de conhecimento, criatividade e experiência necessários, por parte dos participantes envolvidos, para o sucesso. A partir da presente pesquisa, diversas novas aplicações podem ser sugeridas, visando a descoberta de novas informações e fontes de conhecimento possíveis. Algumas propostas de trabalhos futuros seriam:

- Adaptação do método proposto a outras ferramentas colaborativas, como: fóruns, grupos de discussão, *Wikis*, etc.;
- Integração do método proposto com um repositório de conhecimento organizacional, de forma que os Artefatos de Conhecimento trocados pelos Participantes Ativos do processo possam ser automaticamente armazenados;
- Aplicação de novas técnicas de mineração de texto, como *clustering*, de forma a identificar grupos de indivíduos pelos seus papéis no processo, por exemplo;
- Desenvolvimento de novas soluções capazes de extrair padrões, a partir dos textos minerados, que permitam identificar perguntas e respostas frequentes, tal como uma possível sequência de execução entre as atividades;
- Elaboração de notações gráficas específicas para a representação de Processos Intensivos em Conhecimento.

Referências Bibliográficas

ABECKER, A., AITKEN, S., SCHMALHOFER, F. & TSCHAITSCHIAN, B., 1998: "KARATEKIT: Tools for the Knowledge-Creating Company". In Proceedings of the 11th Banff Knowledge Acquisition for Knowledge Based Systems Workshop, v.2, pp.1-18, Banff, Canada.

ABECKER, A., BERNARDI, A., MAUS, H., SINTEK, M., WENZEL, C., 2000, "Information supply for business processes: coupling workflow with document analysis and information retrieval". In *Proceedings of Knowledge-Based Systems*, v.13, pp. 271-284, Amsterdam, Netherlands.

ABECKER, A., BERNARDI, A., DIOUDIS, S., VAN ELST, L., HERTERICH, R., HOUY, C., LEGAL, M., MENTZAS, G. & MÜLLER, S., 2001, "Workflow-Embedded Organizational Memory Access: The DECOR Project". In International Conference on Artificial Intelligence (IJCAI), Workshop on Knowledge Management and Organizational Memories, Seattle.

ABECKER, A., 2004, "*Business-Process Oriented Knowledge Management: Concepts, Methods, and Tools*". Ph.D. thesis, University of Karlsruhe.

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.

BAEZA YATES, R. & RIBEIRO NETO, B., 1999, *Modern Information Retrieval*, Addison-Wesley Longman Publishing Co., Inc., Boston, Massachusetts.

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.

BUSCH, P., 2008, *Tacit Knowledge in Organizational Learning*. Hershey, IGI Publishing.

BUZAN, T. & BUZAN, B., 1996, *The Mind Map Book: How to Use Radiant Thinking to Maximize Your Brain's Untapped Potential*, Plume.

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

CAVALCANTI, M. C., TARGINO, R., BAIÃO, F. *et al.*, 2005, "Managing structural genomic workflows using Web services", *Data and Knowledge Engineering* v.53, 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), Springer Verlag, v.5190, pp. 51-60.

DAVENPORT, T. H., JARVENPAA, S. L., BEERS, M. C., 1996, "Improving Knowledge Work Processes", *Sloan Management Review*, vol. 37, pp. 53-65.

DAVENPORT, T. H. & PRUSAK, L., 2000, *Working Knowledge: How Organizations Manage What They Know*, Harvard Business School Press, Boston, Massachusetts.

DRUCKER, P., 1993, *Sociedade Pós Capitalista*. São Paulo, Pioneira.

EPPLER, M. J., SEIFRIED, P. M., RÖPNACK, A., 1999, "Improving knowledge intensive processes through an enterprise knowledge medium". In Proceedings of the 1999 ACM SIGCPR conference on Computer personnel research, pp. 222-230.

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.

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, p. 611-616, Santiago, Chile.

GONÇALVES, J. C. A. R., 2010, *Story Mining: Elicitação de Processos de Negócio a partir de Group Storytelling e Técnicas de Mineração de Texto*. Dissertação de Mestrado. Universidade Federal do Estado do Rio de Janeiro (UNIRIO), Rio de Janeiro, Brasil.

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), p. 295-300.

GRONAU N., MÜLLER C., USLAR M., 2004, "The KMDL Knowledge Management Approach: Integrating Knowledge Conversions and Business Process Modeling". In Practical Aspects of Knowledge Management, pp.1-11, Springer-Verlag, Berlin, Heidelberg.

GRONAU, N. & WEBER, E., 2004, "Defining an Infrastructure for Knowledge Intensive Business Processes". In *Proceedings of I-Know '04 - 4th International Conference on Knowledge Management*, Graz, Austria.

HAMEL G., PRAHALAD C.K., 1990, "The Core Competence of the Corporation", *Harvard Business Review*, vol. 68, pp. 79-91.

HAMMER, M., 1996, *Beyond Reengineering*. Harper Business, New York.

HEARST, M., 1999, "Untangling Text Data Mining". In Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics (ACL), University of Maryland, USA.

HEISIG, P., 2003, "Wissensmanagement in Industriellen Geschäftsprozessen". Em Alemão. *Industrie Management*, v. 19, n. 3, pp. 22–25, Gito-Verlag, Berlin.

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

KIDD A., 1994, "The Marks are on the Knowledge Worker. Human Factors in Computing Systems". In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 186–191, Boston, Massachusetts.

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

KÜHN, O. & ABECKER, A., 1997, "Corporate Memories for Knowledge Management in Industrial Practice: Prospects and Challenges". *Journal of Universal Computer Science*.

MAURER, H., 1999, "The heart of the Problem: Knowledge Management and Knowledge Transfer". In *Enable'99*, Espoo-Vanta Institute of Technology.

MENTZAS, G., APOSTOLOU, D., YOUNG, R. & ABECKER, A., 2002, "*Knowledge Asset Management - Beyond the Product-centric and the Process-centric Approach*". Springer-Verlag, Berlin, Heidelberg.

NONAKA I., 1994, *A Dynamic Theory of Organizational Knowledge Creation*. *Organization Science*, v. 5, n. 1.

NONAKA, I. & TAKEUCHI, H., 1995, *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*, Oxford University Press, New York.

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

OVUM, 1998, *Knowledge Management: Applications, markets and technologies*. Ovum Ltd.

PAPAVASSILIOU, G., NTIOUDIS, S., ABECKER, A., MENTZAS, G., 2002, "Managing Knowledge in weakly-structured Administrative Processes". In *Proceedings of the Third European Conference on Organizational Knowledge, Learning, and Capabilities (OKLC)*, Athens, Greece.

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.

SCHEIR, P., 2002, *Supporting Customer Relationship Management through Knowledge Management*. Master's Thesis. Institute for Information Systems and Computer Media.

SCHREIBER G., AKKERMANS H., ANJEWIERDEN A., DE HOOG R., SHADBOLD N., VAN DER VELDE W., WIELINDA B., 2000, "Knowledge Engineering and Management, The CommonKADS Methodology". *Artificial Intelligence in Medicine*, v.24, n. 1, pp. 97-102. The MIT Press. Cambridge, Massachusetts.

SCHWARTZ, A. & HEARST, M., 2003, "A Simple Algorithm for Identifying Abbreviation Definitions in Biomedical Text". In Proceedings of the Pacific Symposium on Biocomputing (PSB), pp.451-462, Hawaii, USA.

STEELS, L., 1993, "Corporate knowledge management". In Proceedings of ISMICK'93 Conference, pp. 9-30, Compiègne, France.

UDOH, E. (2007), "Mining E-Mail Content for a Small Enterprise". In Innovations and Advanced Techniques in Computer and Information Sciences and Engineering, pp.179-182, Springer, Netherlands.

VAN DER AALST, W. M., NIKOLOV, A., 2007, "EMailAnalyzer: An E-Mail Mining Plug-in for the ProM Framework", BPM Center Report BPM, pp.7-16, BPMCenter.org.

VAN DER AALST, W. M., WEIJTERS, A., 2005, *Process-Aware Information Systems: Bridging People and Software through Process Technology*, Wiley-Interscience, Hoboken, New Jersey.

VON HAGEN, C. R., RATZ, D., POVALEJ, R., 2005, "Evolution of Petri-net Modeled Knowledge Intensive Processes". In ISTA, v.63, pp. 200-203.

WEBER, R., AHA, D.W., BECERRA FERNANDEZ, I., 2001, "Intelligent Lessons Learned Systems". *International Journal of Expert Systems Research & Applications*

WENGER, E., 1998, *Communities of Practice. Learning, Meaning and Identity*. Cambridge University Press.

Apêndice A – Estrutura de Metadados

Neste apêndice estão presentes todas as estruturas de metadados mencionadas no Capítulo Quatro referente à proposta de solução. Estas estruturas consistem nos arquivos XSD gerados para representar os arquivos XML a serem gerados pelo método proposto.

Arquivo XSD Gerado para a Fase de Extração

```
<?xml version="1.0" encoding="utf-8"?>
<xs:schema id="Extracao" xmlns="" xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="Extracao" msdata:IsDataSet="true" msdata:Locale="en-US">
    <xs:complexType>
      <xs:choice minOccurs="0" maxOccurs="unbounded">
        <xs:element name="Email">
          <xs:complexType>
            <xs:sequence>
              <xs:element name="DataEnvio" type="xs:string" minOccurs="0" />
              <xs:element name="Remetente" type="xs:string" minOccurs="0" />
              <xs:element name="Destinatario" nillable="true" minOccurs="0" maxOccurs="unbounded">
                <xs:complexType>
                  <xs:simpleContent msdata:ColumnName="Destinatario_Text" msdata:Ordinal="0">
                    <xs:extension base="xs:string">
                    </xs:extension>
                  </xs:simpleContent>
                </xs:complexType>
              </xs:element>
              <xs:element name="Copiado" nillable="true" minOccurs="0" maxOccurs="unbounded">
                <xs:complexType>
                  <xs:simpleContent msdata:ColumnName="Copiado_Text" msdata:Ordinal="0">
                    <xs:extension base="xs:string">
                    </xs:extension>
                  </xs:simpleContent>
                </xs:complexType>
              </xs:element>
              <xs:element name="ArquivoAnexo" nillable="true" minOccurs="0" maxOccurs="unbounded">
                <xs:complexType>
                  <xs:simpleContent msdata:ColumnName="ArquivoAnexo_Text" msdata:Ordinal="0">
                    <xs:extension base="xs:string">
                    </xs:extension>
                  </xs:simpleContent>
                </xs:complexType>
              </xs:element>
              <xs:element name="Assunto" type="xs:string" minOccurs="0" />
              <xs:element name="Mensagem" type="xs:string" minOccurs="0" />
            </xs:sequence>
          </xs:complexType>
        </xs:element>
      </xs:choice>
    </xs:complexType>
  </xs:element>
</xs:schema>
```

Arquivo XSD Gerado para a Fase de Mineração de Textos

```
<?xml version="1.0" encoding="utf-8"?>
<xs:schema id="MineracaoTexto" xmlns="" xmlns:xs="http://www.w3.org/2001/XMLSchema" xmlns:msdata
-="urn:schemas-microsoft-com:xml-msdata">
  <xs:element name="MineracaoTexto" msdata:IsDataSet="true" msdata:Locale="en-US">
    <xs:complexType>
      <xs:choice minOccurs="0" maxOccurs="unbounded">
        <xs:element name="Email">
          <xs:complexType>
            <xs:sequence>
              <xs:element name="DataEnvio" type="xs:string" minOccurs="0" />
              <xs:element name="Remetente" type="xs:string" minOccurs="0" />
              <xs:element name="Assunto" type="xs:string" minOccurs="0" />
              <xs:element name="Destinatario" nillable="true" minOccurs="0" maxOccurs="
unbounded">
                <xs:complexType>
                  <xs:simpleContent msdata:ColumnName="Destinatario_Text" msdata:Ordinal="0">
                    <xs:extension base="xs:string">
                      </xs:extension>
                    </xs:simpleContent>
                  </xs:complexType>
                </xs:element>
              <xs:element name="Copiado" nillable="true" minOccurs="0" maxOccurs="unbounded">
                <xs:complexType>
                  <xs:simpleContent msdata:ColumnName="Copiado_Text" msdata:Ordinal="0">
                    <xs:extension base="xs:string">
                      </xs:extension>
                    </xs:simpleContent>
                  </xs:complexType>
                </xs:element>
              <xs:element name="ArquivoAnexo" nillable="true" minOccurs="0" maxOccurs="
unbounded">
                <xs:complexType>
                  <xs:simpleContent msdata:ColumnName="ArquivoAnexo_Text" msdata:Ordinal="0">
                    <xs:extension base="xs:string">
                      </xs:extension>
                    </xs:simpleContent>
                  </xs:complexType>
                </xs:element>
              <xs:element name="Atividades" minOccurs="0" maxOccurs="unbounded">
                <xs:complexType>
                  <xs:sequence>
                    <xs:element name="Atividade" minOccurs="0" maxOccurs="unbounded">
                      <xs:complexType>
                        <xs:sequence>
                          <xs:element name="Ator" type="xs:string" minOccurs="0" />
                          <xs:element name="Acao" type="xs:string" minOccurs="0" />
                          <xs:element name="Parametro" type="xs:string" minOccurs="0" />
                        </xs:sequence>
                      </xs:complexType>
                    </xs:element>
                  </xs:sequence>
                </xs:complexType>
              </xs:element>
            </xs:sequence>
          </xs:complexType>
        </xs:choice>
      </xs:complexType>
    </xs:element>
  </xs:schema>
```

Arquivo XSD do *FreeMind* Utilizado para a Fase de Representação

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema" elementFormDefault="qualified">

  <xs:element name="Parameters">
    <xs:complexType>
      <!-- Is the time management plugin.-->
      <xs:attribute name="REMINDUSERAT" type="xs:integer" use="optional"/>
    </xs:complexType>
  </xs:element>
  <!-- Used for node notes.-->
  <xs:element name="text">
    <xs:complexType/>
  </xs:element>

  <xs:element name="arrowlink">
    <xs:complexType>
      <xs:attribute name="COLOR" type="xs:string" use="optional"/>
      <xs:attribute name="DESTINATION" type="xs:string" use="required"/>
      <xs:attribute name="ENDARROW" type="xs:string" use="optional"/>
      <xs:attribute name="ENDINCLINATION" type="xs:string" use="optional"/>
      <xs:attribute name="ID" type="xs:string" use="optional"/>
      <xs:attribute name="STARTARROW" type="xs:string" use="optional"/>
      <xs:attribute name="STARTINCLINATION" type="xs:string" use="optional"/>
    </xs:complexType>
  </xs:element>

  <xs:element name="cloud">
    <xs:complexType>
      <xs:attribute name="COLOR" type="xs:string" use="optional"/>
    </xs:complexType>
  </xs:element>

  <xs:element name="edge">
    <xs:complexType>
      <xs:attribute name="COLOR" type="xs:string" use="optional"/>
      <xs:attribute name="STYLE" type="xs:string" use="optional"/>
      <xs:attribute name="WIDTH" type="xs:string" use="optional"/>
    </xs:complexType>
  </xs:element>

  <xs:element name="font">
    <xs:complexType>
      <xs:attribute name="BOLD" use="optional">
        <xs:simpleType>
          <xs:restriction base="xs:string">
            <xs:enumeration value="true"/>
          </xs:restriction>
        </xs:simpleType>
      </xs:attribute>
      <xs:attribute name="ITALIC" use="optional">
        <xs:simpleType>
          <xs:restriction base="xs:string">
            <xs:enumeration value="true"/>
          </xs:restriction>
        </xs:simpleType>
      </xs:attribute>
    </xs:complexType>
  </xs:element>
</xs:schema>
```

```

    <xs:enumeration value-'false'/'>
  </xs:restriction>
</xs:simpleType>
</xs:attribute>
<xs:attribute name-'NAME' type-'xs:string' use-'required'/'>
<xs:attribute name-'SIZE' use-'required' type-'xs:integer'/'>
</xs:complexType>
</xs:element>

<xs:element name-'hook'>
  <xs:complexType>
    <xs:sequence>
      <xs:element ref-'Parameters' minOccurs-'0' maxOccurs-'1'/'>
      <xs:element ref-'text' minOccurs-'0' maxOccurs-'1'/'>
    </xs:sequence>
    <xs:attribute name-'NAME' type-'xs:string' use-'required'/'>
  </xs:complexType>
</xs:element>

<xs:element name-'icon'>
  <xs:complexType>
    <xs:attribute name-'BUILTIN' type-'xs:string' use-'required'/'>
  </xs:complexType>
</xs:element>

<xs:element name-'map'>
  <xs:complexType>
    <xs:sequence>
      <xs:element ref-'node'/'>
    </xs:sequence>
    <xs:attribute name-'version' type-'xs:string' use-'required'/'>
  </xs:complexType>
</xs:element>

<xs:element name-'node'>
  <xs:complexType>
    <xs:choice minOccurs-'0' maxOccurs-'unbounded'>
      <xs:element ref-'arrowlink'/'>
      <xs:element ref-'cloud'/'>
      <xs:element ref-'edge'/'>
      <xs:element ref-'font'/'>
      <xs:element ref-'hook'/'>
      <xs:element ref-'icon'/'>
      <xs:element ref-'node'/'>
    </xs:choice>
    <xs:attribute name-'BACKGROUND_COLOR' type-'xs:string' use-'optional'/'>
    <xs:attribute name-'COLOR' type-'xs:string' use-'optional'/'>
    <xs:attribute name-'FOLDED' use-'optional'>
    <xs:simpleType>
      <xs:restriction base-'xs:string'>
        <xs:enumeration value-'true'/'>
        <xs:enumeration value-'false'/'>
      </xs:restriction>
    </xs:simpleType>
  </xs:attribute>
  <xs:attribute name-'ID' type-'xs:ID' use-'optional'/'>
  <xs:attribute name-'LINK' type-'xs:string' use-'optional'/'>
  <xs:attribute name-'POSITION' use-'optional'>
  <xs:simpleType>
    <xs:restriction base-'xs:string'>
      <xs:enumeration value-'left'/'>
      <xs:enumeration value-'right'/'>
    </xs:restriction>
  </xs:simpleType>
  </xs:attribute>
  <xs:attribute name-'STYLE' type-'xs:string' use-'optional'/'>
  <xs:attribute name-'TEXT' type-'xs:string' use-'required'/'>
  <xs:attribute name-'CREATED' type-'xs:integer' use-'optional'/'>
  <xs:attribute name-'MODIFIED' type-'xs:integer' use-'optional'/'>
  <xs:attribute name-'HGAP' type-'xs:integer' use-'optional'/'>
  <xs:attribute name-'VGAP' type-'xs:integer' use-'optional'/'>
  <xs:attribute name-'VSHIFT' type-'xs:integer' use-'optional'/'>
  <xs:attribute name-'ENCRYPTED_CONTENT' type-'xs:string' use-'optional'/'>
</xs:complexType>
</xs:element>
</xs:schema>

```

Apêndice B – Resultado do Estudo de Caso

Neste apêndice, encontra-se ilustrada uma visão ampliada do resultado gerado pelo método proposto por esta pesquisa após sua aplicação ao Estudo de Caso apresentado. Para uma análise detalhada, é possível navegar pelo Mapa Mental gerado, contido no CD anexo a esta dissertação.

