

CENTRO UNIVERSITÁRIO EURÍPIDES DE MARÍLIA

FUNDAÇÃO DE ENSINO “EURÍPIDES SOARES DA ROCHA”

BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

**DESENVOLVIMENTO DE ABORDAGEM DE FUSÃO
DE INFORMAÇÕES UTILIZANDO TÉCNICAS
SEMÂNTICAS PARA A MELHORIA DA QUALIDADE
DA INFORMAÇÃO DE SITUAÇÕES DE RISCO**

VALDIR AMANCIO PEREIRA JUNIOR

ORIENTADOR: PROF. DR. LEONARDO CASTRO BOTEGA

Marília - SP
Novembro/2017

CENTRO UNIVERSITÁRIO EURÍPIDES DE MARÍLIA

FUNDAÇÃO DE ENSINO "EURÍPIDES SOARES DA ROCHA"

BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

**DESENVOLVIMENTO DE ABORDAGEM DE FUSÃO
DE INFORMAÇÕES UTILIZANDO TÉCNICAS
SEMÂNTICAS PARA A MELHORIA DA QUALIDADE
DA INFORMAÇÃO DE SITUAÇÕES DE RISCO**

VALDIR AMANCIO PEREIRA JUNIOR

Monografia apresentada ao Centro
Universitário Eurípides de Marília como parte
dos requisitos necessários para a obtenção do
grau de Bacharel em Ciência da Computação.
Orientador: Prof. Dr. Leonardo Castro Botega

Marília - SP

Novembro/2017



CENTRO UNIVERSITÁRIO EURÍPIDES DE MARÍLIA - UNIVEM
MANTIDO PELA FUNDAÇÃO DE ENSINO "EURÍPIDES SOARES DA ROCHA"

BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

Valdir Amancio Pereira Junior

DESENVOLVIMENTO DE ABORDAGEM DE FUSÃO DE INFORMAÇÕES
UTILIZANDO TÉCNICAS SEMÂNTICAS PARA A MELHORIA DA QUALIDADE
DA INFORMAÇÃO DE SITUAÇÕES DE RISCO

Banca examinadora da monografia apresentada ao Curso de Bacharelado em
Ciência da Computação do UNIVEM/F.E.E.S.R., para obtenção do Título de
Bacharel em Ciência da Computação.

Nota: 10.0 (DEZ)

Orientador: Leonardo Castro Botega *Le C.*

1º.Examinador: Fabio Piola Navarro *F. P. N.*

2º.Examinador: Guilherme Rodrigues Bilar *Guilherme R. Bilar*

Marília, 27 de novembro de 2017.

Dedico este trabalho e toda minha graduação a minha família, de maneira especial aos meus pais, por todo apoio, confiança, carinho, sacrifícios e conselhos, me inspirando a cada momento durante os últimos quatro anos, só cheguei até aqui por causa vocês. Di e Li, eu não seria nada sem vocês meus eternos mestres. Família vai muito além de laços de sangue ou paternidade.

AGRADECIMENTO

Agradeço primeiramente ao Centro Universitário Eurípides de Marília pela bolsa de estudos cedida e todas as demais oportunidades oferecidas, sem vocês não teria usufruído de todas as ótimas experiências durante a graduação.

Agradeço aos meus pais, Valdir e Elisângela, aos meus irmãos, André, Pedro e Chester, por cada instante vivido e que me ajudaram com os mais diversos assuntos, por cada conselho e risada.

Agradeço meu orientador, conselheiro e sócio Dr. Leonardo Castro Botega, por cada ensinamento transmitido, desde antes da graduação e por cada conselho dado, por me propiciar as melhores oportunidades e pela sua confiança. Aos meus amigos do Grupo de Interação Humano-Computador e do COMPSI, por cada apoio durante as semanas mais difíceis do curso, por cada risada a tarde no 64 e entre as aulas, especialmente Mayumi, Thiago, Jordan, Gustavo, Jordana, Zanco, Gabriel e Cássio.

A todo corpo docente da instituição, especialmente Ms. Rodolfo Barros Chiaramonte, Dr. Elvis Fusco, Dr. Fabio Dacencio e Dr. Allan Oliveira.

Nós só podemos ver um pouco do futuro, mas o suficiente para perceber que há muito a fazer.

Alan Turing

RESUMO

Consciência Situacional (Situational Awareness – SAW) se refere ao nível de consciência que um indivíduo ou equipe detém sobre uma situação. No domínio de gerenciamento de riscos, falhas de SAW podem induzir operadores humanos a erros no processo decisório e acarretar riscos à vida, ao patrimônio ou ao meio ambiente. Processos de fusão de dados apresentam oportunidades para melhorar a SAW de operadores de sistemas de avaliação de situações de risco. Entretanto, cenários dinâmicos estão sujeitos a problemas de qualidade das informações, especialmente quando os dados são provenientes de Inteligência Humana (HUMINT). O estado da arte em técnicas de fusão de dados e informações apresenta abordagens com limitado emprego da qualidade da informação no processo de fusão. Além disso, as soluções são restritas a mecanismos sintáticos para a determinação de sinergia entre informações, influenciando negativamente a qualidade da informação dos resultados da fusão. Assim, esta proposta objetiva o desenvolvimento de um novo processo de fusão semântica de informações, mais especificamente uma evolução do modelo Quantify (*Quality-aware Human-Driven Information Fusion Model*), que tem como principal característica o emprego contínuo da avaliação sintática da informação ao longo do processo de fusão, além de empregar a qualidade da informação como forma de parametrização de suas rotinas. Nesta nova abordagem, uma rotina de fusão semântica foi acrescentada ao modelo, aumentando consideravelmente seu poder de inferência. Resultados demonstraram-se promissores, validado por um estudo de caso com informações reais de situações de risco envolvendo dados criminais.

Palavras-chave: Fusão Semântica, Fusão de Informações, Consciência Situacional, Gerenciamento de Riscos.

ABSTRACT

Situational Awareness (SAW) refers to the level of consciousness that an individual or team holds about a situation. In the domain of risk management, SAW failures can induce human operators to make mistakes in the decision-making process and to risk life, property or the environment. Data fusion processes present opportunities to improve the SAW of operators of risk assessment systems. However, dynamic scenarios are subject to information quality problems, especially when the data are derived from human intelligence (HUMINT). The state of the art in data and information fusion techniques present approaches with limited employment of quality information in the fusion process. Furthermore, solutions are restricted to syntactic mechanisms for determining synergy between information, negatively influencing the quality of the information of the results of the fusion. Thus, this proposal aims to develop a new process of semantic information fusion, through an implementation aligned with the model Quantify (Quality-aware Human-Driven Information Fusion Model), whose main characteristic is the continuous employment of the syntactic evaluation of information throughout the fusion process, besides using quality of the information as a way of parameterizing these. In this new approach, a semantic fusion routine was added to the model, greatly increasing its inference capacity. Results have shown to be promising, validated by a case study with real information of risk situations involving criminal data.

Keywords: Semantic Fusion, Information Fusion, Situational Awareness, Risk Management.

LISTA DE FIGURAS

Figura 2.1: Modelo de Consciência Situacional de Endsley (Traduzido e Adaptado de Endsley, 1988)	22
Figura 3.1: Modelo JDL de Fusão de Dados. (Traduzido e Adaptado de Hall e MacMullen, 2004).....	32
Figura 3.2: Exemplo de Ontologia	39
Figura 5.1: Parte da Análise de Tarefas Dirigida a Objetivos Gerado a Partir dos Resultados Obtidos no Questionário Encaminhado à PMESP	47
Figura 5.2: Modelo Quantify com Destaque para o Módulo de Fusão (Adaptado Botega, 2015).....	48
Figura 5.3: Processo de Fusão Sintática implementado para identificação e classificação de termos presentes em registros criminais.....	49
Figura 5.4: Pseudocódigo demonstrando o processo inicial de Fusão Sintática.....	50
Figura 5.5: Pseudocódigo ilustrando o núcleo da Fusão Sintática de Informações ..	52
Figura 5.6: Etapas da Metodologia 101 (Traduzido e Adaptado de Noy & McGuinness, 2001)	54
Figura 5.7: Processo Integrado de Fusão de Informações contando com Camada Sintática e Semântica (Adaptado de Junior, 2016)	56
Figura 5.8: Exemplo de Saída dos Módulos de Aquisição e Preparação de Dados do Novo Processo Integrado de Fusão	58
Figura 5.9: Resultado do Processo de Avaliação de Objetos realizado com avaliação Sintática	60
Figura 5.10: Exemplo de Saída com Índices de Qualidade Locais e Global	60
Figura 5.11: Pseudocódigo demonstrando a base da Busca de Informações Sinérgicas.....	62
Figura 5.12: Demonstração de Resultado de uma Busca de Informações Sinérgicas	62
Figura 5.13: Pseudocódigo Demonstrando o Core da Associação Multicritério	63
Figura 5.14: Ontologia Desenvolvida para o Domínio de Gerenciamento de Riscos (Roubo e Furto).....	66
Figura 5.15: Exemplo de Consulta SPARQL para Inferir as Propriedades Sobre Determinadas Classes	68
Figura 5.16: Resultado da Consulta SPARQL Realizada na Figura Anterior	68

Figura 5.17: Exemplo de Resultado da Avaliação de Objetos com Abordagem Semântica	68
Figura 5.18: Exemplo de Representação de Índices de Qualidade Utilizando DQV .	69
Figura 5.19: Demonstração de Resultado de uma Busca de Informações Sinérgicas através da Semântica.....	70
Figura 5.20: Pseudocódigo de Integração Semântica de Informações	71
Figura 6.1: Denúncia 1 pós Avaliação de Objetos e persistida na Camada de Representação da Informação como Situação e com Índices de Qualidade	75
Figura 6.2: Denúncia 2 pós Avaliação de Objetos e persistida na Camada de Representação da Informação como Situação e com Índices de Qualidade.....	76
Figura 6.3: Denúncia 3 pós Avaliação de Objetos e persistida na Camada de Representação da Informação como Situação e com Índices de Qualidade	77
Figura 6.4: Situação Obtida com a Fusão Sintática Durante o Estudo de Caso	78
Figura 6.5: RDF da Situação Gerada a Partir da Avaliação de Objetos da Denúncia 1	80
Figura 6.6: RDF da Situação Gerada a Partir da Avaliação de Objetos da Denúncia 2	81
Figura 6.7: RDF da Situação Gerada a Partir da Avaliação de Objetos da Denúncia 3	81
Figura 6.8: Resultado do Processo Semântico de Busca e Integração de Informações Sinérgicas com Informações do Estudo de Caso	83
Figura 6.9: Resultado do Processo Semântico de Busca e Integração de Informações Sinérgicas Aplicado o DQV para Representar a Qualidade	84

LISTA DE TABELAS

Tabela 5.1: Tabela de Classes da Ontologia, descrevendo o que cada classe presente representa para a situação.....	64
--	----

LISTA DE ABREVIATURAS E SIGLAS

DQV - *Data Quality Vocabulary*

GDTA - *Goal-Directed Task Analysis*

HUMINT - *Human Intelligence*

JDL - *Joint Directors Laboratories*

JSON - *JavaScript Object Notation*

JSON-LD - *JavaScript Object Notation for Linked Data*

KNN - *K-Nearest Neighbors*

OWL - *Web Ontology Language*

PMESP - *Polícia Militar do Estado de São Paulo*

RDF - *Resource Description Framework*

SAW - *Situational Awareness*

SPARQL - *Simple Protocol and RDF Query Language*

SQL - *Structured Query Language*

W3C - *World Wide Web Consortium*

SUMÁRIO

CAPÍTULO 1 - INTRODUÇÃO.....	13
1.1 Contexto.....	13
1.2 Motivação.....	14
1.3 Problemática de Pesquisa.....	15
1.4 Objetivos	15
1.5 Metodologia de Desenvolvimento do Trabalho	16
1.6 Organização do Trabalho.....	17
CAPÍTULO 2 - CONSCIÊNCIA SITUACIONAL E SISTEMAS CRÍTICOS.....	19
2.1 Consciência Situacional	19
2.2 Modelagem da Consciência Situacional.....	20
2.2.1 Fatores Influentes Durante o Processo	22
2.3 Aplicação e Domínios Críticos de Consciência Situacional.....	24
2.4 Mecanismos de Apoio à Consciência Situacional	26
2.5 Considerações Finais.....	27
CAPÍTULO 3 - FUSÃO DE DADOS E INFORMAÇÕES APLICADO AO DOMÍNIO DE AVALIAÇÃO DE SITUAÇÕES DE RISCO	29
3.1 Fusão de Dados e Informações	29
3.2 Modelo JDL de Fusão de Dados.....	31
3.2.1 A Evolução do Modelo JDL	32
3.2.2 Modelo JDL e o Modelo de Endsley	35
3.3 Fusão de Dados e Informações e a Participação Humana	35
3.4 Adequação dos Métodos de Fusão para Suportar Informações em Alto-Nível ...	37
3.5 Considerações Finais.....	40
CAPÍTULO 4 - ESTADO DA ARTE	41
CAPÍTULO 5 - DESENVOLVIMENTO DE ABORDAGEM DE FUSÃO DE INFORMAÇÕES UTILIZANDO TÉCNICAS SEMÂNTICAS PARA A MELHORIA DA QUALIDADE DA INFORMAÇÃO DE SITUAÇÕES DE RISCO	44
5.1 O Domínio de Gerenciamento de Risco.....	44

5.2 Levantamento de Requisitos	46
5.3 Desenvolvimento do Processo Integrado de Fusão de Informações	46
5.3.1 Processo de Fusão Sintática de Informações	48
5.3.1.1 Estabelecendo os Limites da Fusão Sintática	52
5.3.2 Desenvolvimento de uma Ontologia para Representar o Conhecimento do Domínio	54
5.3.3 Novo Processo Integrado de Fusão Sintática e Semântica.....	56
5.3.3.1 Aquisição e Preparação dos Dados	57
5.3.3.2 Módulo de Avaliação de Qualidade de Dados e Informações	58
5.3.3.3 Um Novo Processo de Fusão Sintática Utilizando Qualidade	58
5.3.3.4 Processo de Fusão Semântica utilizando Qualidade	64
5.4 Considerações Finais	72
CAPÍTULO 6 - ESTUDO DE CASO COM O PROCESSO INTEGRADO DE FUSÃO DE INFORMAÇÕES	73
CAPÍTULO 7 - CONCLUSÕES.....	85
7.1 Publicações Obtidas.....	86
REFERÊNCIAS.....	89

Capítulo 1

INTRODUÇÃO

Neste Capítulo é apresentada uma introdução sobre os temas que estruturam este trabalho, bem como a motivação e os objetivos, além da metodologia adotada para o desenvolvimento do trabalho e como este está estruturado.

1.1 Contexto

Consciência Situacional (SAW) é um processo cognitivo importante para os tomadores de decisão em diversas áreas críticas e diz respeito à percepção da presença de entidades de interesse em um ambiente, à compreensão do significado de suas ações individuais e coletivas no espaço-tempo e à projeção de seus status em um futuro próximo (Kokar e Endsley, 2012). No domínio de gerenciamento de riscos com base em dados criminais, SAW é um fator crucial para revelar tendências, incidência de ameaças e o aumento ou diminuição de riscos iminentes.

Uma limitada SAW pode comprometer a compreensão de analistas ao que realmente aconteceu e está acontecendo. Levando a uma má tomada de decisão, o que pode resultar em consequências desastrosas para as pessoas, patrimônios ou meio ambiente (Rogova e Bossé, 2010).

Embora não possa garantir decisões de melhor qualidade devido aos fatores humanos particulares, uma melhor SAW pode ajudar analistas de segurança, sujeitos a incertezas, a manter um conhecimento superior sobre eventos criminais passados, situações em curso e atividades ilícitas futuras. Para tais profissionais, adquirir, manter e retomar SAW é um desafio, considerando o grande volume e a

heterogeneidade dos dados criminais existentes no Brasil, e também fundamental para o planejamento de atividades econômicas e de segurança do setor público e privado.

Adicionalmente, o processo de aquisição de SAW é ainda mais desafiador pois em sua maioria os dados são fornecidos por inteligência humana (*Human Intelligence* - HUMINT). Como é o caso das denúncias de crimes relatados às centrais de atendimento das forças de segurança. Tipicamente, os dados HUMINT são incompletos, desatualizados, inconsistentes e às vezes até mesmo influenciado por fatores culturais, influenciando os processos computacionais que processam os dados e estimulam a SAW do analista.

Neste contexto, foi desenvolvido um processo Fusão de Dados e Informações para identificar sinergia entre dados e informações heterogêneas e integra-las, fornecendo informações em menor dimensão, porém mais significativas, visando reduzir incertezas (Kokar e Endsley, 2012) (Batini *et al*, 2009) (Blasch *et al*, 2012) (Laskey *et al*, 2012) (Noy & McGuinness, 2001) (Botega *et al*, 2016) (Carvalho *et al*, 2013). Entretanto, particularidades dos desafios citados exigem processos mais robustos de mineração e integração de elementos relevantes presentes em dados fornecidos por humanos.

Uma das contribuições deste trabalho é mostrar na prática como um processo típico de fusão de dados pode ser aplicado para revelar entidades presentes em dados HUMINT, mais especificamente provenientes do domínio criminal, apresentando as dificuldades e limites de inferência quando utilizadas apenas as referências sintáticas.

1.2 Motivação

Adquirir SAW para apoiar a tomada de decisão em ambientes críticos é um processo desafiador, não apenas devido à dinamicidade dos ambientes e das variáveis de interesse, mas também pela necessidade de ferramentas e mecanismos de suporte, tais como os sistemas de avaliação de situações de risco.

Fatores humanos associados ao operador de tais sistemas podem influenciar negativamente o processo de análise de uma situação, bem como o volume de

informação disponível para que o mesmo obtenha conclusões sobre o que está ocorrendo no ambiente. Entretanto, uma grande quantidade de informação não garante decisões de qualidade, sendo necessário mecanismos para a qualificação da informação empregada em suporte a SAW.

1.3 Problemática de Pesquisa

Técnicas de fusão de dados podem contribuir para a melhoria da informação coletada, processada e representada por sistemas de avaliação de situações. Contudo, os mecanismos atualmente empregados são limitados a integrar superficialmente as informações, com base em parâmetros léxicos e sintáticos de seus atributos, o que pode ser um fator limitante em domínios complexos como o de gerenciamento de riscos. Adicionalmente, as abordagens semânticas conhecidas ainda são limitadas quanto à sua capacidade de determinação de sinergia e integração entre as informações, sendo necessário que sua parametrização seja enriquecida com mecanismos como o proposto por este projeto.

1.4 Objetivos

O objetivo geral da presente proposta é o desenvolvimento de um processo de fusão de informações para a melhoria da qualidade da informação, utilizada para contribuir para a melhoria da consciência situacional de operadores humanos de sistemas de avaliação de situações de risco. Tal processo deve operar integrado e de acordo com as diretrizes do modelo Quantify de fusão de informações, no qual a avaliação constante da qualidade da informação inferida ou transformada é um de seus pilares (Botega *et al*, 2014) (Botega, 2016).

Como objetivo específico, busca-se implementar um mecanismo de busca sinérgica e integração de informações utilizando múltiplos critérios hierárquicos, considerando a variedade de classes e atributos inerentes às entidades de interesse e as diversas possibilidades de associação, do nível mais alto de abstração (ex:

objetos da classe “vítimas” em conjunto com outros objetos semelhantes e as atividades que exercem), até o nível de atributos de tais objetos (ex: características semelhantes que descrevem a vítima, como a sua condição física).

Adicionalmente, será considerada a qualidade da informação como um destes atributos dos objetos, sob dimensões como a completude, atualidade, consistência e relevância. Desta maneira, busca-se associar e integrar informações dentro de um limiar de qualidade de dados de entrada relevantes para o sistema e/ou operador humano, além de estabelecer um padrão desejável de qualidade.

Para operacionalizar tal busca e integração sinérgica entre objetos e atributos, as informações reportadas ao serviço de atendimento a riscos serão analisadas sob as perspectivas sintática e semântica. Em busca de similaridades pela forma como as entidades e seus atributos são descritos e também considerando o que representam (busca por significado sinérgico).

1.5 Metodologia de Desenvolvimento do Trabalho

Esta pesquisa tem natureza exploratória, experimental, qualitativa e dirigida por um estudo de caso. A mesma segue os métodos de desenvolvimento e definições do modelo Quantify de fusão de informações (Botega, 2016). Para que o trabalho atingisse os objetivos propostos, este teve sua execução dividida em etapas complementares que guiaram a sua execução, sendo estas:

1. Pesquisa sobre modelos e algoritmos de fusão de dados e estudo do domínio criminal para compreensão dos dados e prováveis associações sintáticas e semânticas;
2. Levantamento de requisitos necessário do domínio aplicando a Análise de Tarefas Dirigida a Objetivos (*Goal-Directed Task Analysis - GDTA*) nos operadores de sistemas de gerenciamentos de riscos;
3. Levantamento do estado da arte em Fusão Sintática de informações envolvendo SAW, qualidade da Informação ou ontologias, visando pontuar os avanços propostos nesta abordagem;

4. Ensaios e implementações de técnicas de Fusão Sintática com o objetivo de determinar qual melhor conjunto de técnica se adequaria as necessidades da proposta, inspirado pelos dados do domínio criminal;
5. Implementação definitiva de uma fusão sintática junto a testes em ambiente controlado visando provar suas falhas, limites e contribuições oferecidas;
6. Pesquisa de trabalhos correlatos à fusão de informação semânticas e como estes se posicionam em relação à teoria e prática de informações semânticas e com índices de qualidade de informações;
7. Levantamento e análise de técnicas ligadas a gerenciamento e manipulação de dados semânticos para serem aplicados no projeto e demais requisitos da adesão da semântica durante o processo de fusão;
8. Desenvolvimento de um processo que suportasse a aplicação de ambos os processos de fusão, sintático e semântico, assim como a avaliação de qualidade e tratamento de informações;
9. Construção de uma ontologia que capaz de representar os conhecimentos do domínio de maneira computacional tornando possível a implementação de técnicas de Fusão Semântica com seu real potencial e assertividade;
10. Ensaios com algoritmos e técnicas semânticas utilizando a arquitetura e ontologia desenvolvidas para suprir a necessidade do domínio, visando validar e selecionar a técnica que mais se adequa.

Os dados a serem utilizados para a execução das etapas e métodos desta proposta foram primeiramente Boletins de Ocorrência (BOs), fornecidos pela Secretaria de Segurança Pública de São Paulo, mediante lei de acesso à informação (referência da lei).

1.6 Organização do Trabalho

O presente trabalho está organizado da seguinte forma: O Capítulo 2 demonstra e relaciona questões sobre SAW e sua relação com Situações de Risco.

O Capítulo 3 apresenta as técnicas de Fusão de Dados e Informações e as aplicações no domínio em questão. O Capítulo 4 dedica-se ao desenvolvimento deste trabalho, com o processo de fusão, processos internos e técnicas desenvolvimento para satisfazer a proposta do mesmo. No Capítulo 5 é realizado um Estudo de Caso como demonstração da eficácia e comprovando a necessidade desta abordagem no domínio. Finalmente, o Capítulo 6 traz as conclusões obtidas com este trabalho.

Capítulo 2

CONSCIÊNCIA SITUACIONAL E SISTEMAS CRÍTICOS

Este capítulo abordará uma das bases desse trabalho, apresentando conceitos, modelos, fatores que envolvem o processo cognitivo de Consciência Situacional (*Situational Awareness – SAW*) e principalmente como este processo acontece e evolui em situação de risco.

2.1 Consciência Situacional

A Consciência Situacional é um processo cognitivo humano que representa a capacidade de observar, interpretar e entender o ambiente ou situação em um determinado momento. Segundo Endsley (1988) seria a “percepção de elementos no ambiente em um volume de tempo e espaço, a compreensão do seu significado e projeção de seu estado em um futuro próximo”.

SAW está presente em quase todas as ações realizadas por humanos, sendo observada desde tempos remotos onde as decisões para sobrevivência de uma comunidade ou humano, como caça e locomoção, eram tomadas observando o ambiente e tempo (Endsley, 1988).

Presente nas ações mais simples e corriqueiras executadas até as operacionais e táticas, que exigem um nível de consciência maior, por exemplo, um jogador em uma partida de futebol precisa analisar as condições do campo e disposição dos demais jogadores antes de realizar uma jogada. Sendo que para

cada tarefa ou objetivo as informações envolvidas devem ser observadas de formas diferentes conforme as razões que envolvem o humano (Stanton *et al*, 2001) (Kokar e Endsley, 2012).

Um dos primeiros domínios a aprofundar os conhecimentos e adotar o conceito de Consciência Situacional foi o militar, por sua demanda de altíssimos níveis de SAW em situações críticas, estando presentes desde os primórdios da aeronáutica. Todavia esse conceito não fica restrito a este domínio, sendo aplicado em outros tão importantes quanto, como controle de tráfego aéreo, diagnóstico médico, análise de mercado financeiro, monitoramento de desastres e outros, principalmente aqueles críticos ao tempo e que envolvem muitos fatores, como vida e patrimônio (Endsley, 2001) (Kokar e Endsley, 2012).

Apesar das diferenças práticas para cada um desses domínios, temos um ponto em comum para todos, são ambientes dinâmicos que tem seus estados alterados independente da ação de um agente, seja humano ou máquina. Essa característica faz que SAW seja um processo contínuo e incremental, exigindo atenção total do humano para que possa estar realmente ciente de todo ambiente e poder realizar ações baseadas no conhecimento que construiu.

Partindo dessas considerações, podemos listar SAW como um dos pilares em tomadas de decisão. Entretanto, ter Consciência Situacional não é o suficiente para construir e suportar uma boa decisão, considerando ambiente dinâmicos e as respectivas variáveis envolvidas. Assim, para que haja maior probabilidade de uma tomada de decisão eficaz e concisa com uma situação, é necessário melhorar de forma continua a SAW, isso para a maioria das situações (Endsley e Jones, 2012) (Botega, 2015).

2.2 Modelagem da Consciência Situacional

Dentre todos as abordagens conceituais em torno da Consciência Situacional, como está empregada na vida dos humanos, podendo ser um processo corriqueiro em que não se percebe que está processando o tomando decisões baseadas em fatores do ambiente, desde aquele que o humano está focado em uma situação

específica para fortalecer sua compreensão e tomar embasar da melhor maneira possível sua decisão final.

Existe a necessidade de estudar a fundamentar como este processo ocorre e o que é realmente considerando, para isso alguns modelos para SAW foram propostos, cada um sobre uma perspectiva de aquisição e utilização das informações presentes no ambiente. A solução proposta por este trabalho terá como base o modelo proposto por Endsley (1988), em que a SAW é guiada pela aquisição e uso de informações, tendo esta como um processamento sobre demanda.

Vendo a Consciência Situacional como um processo e com elementos envolvidos, Endsley (2001) distribui e relaciona estes em três níveis, descritos a baixo, estando passíveis a agentes externos, como tarefas e sistemas.

- Nível um: Percepção dos elementos, estados, atributos e a dinâmica dos elementos que compõem o ambiente. Este primeiro passo normalmente de forma visual, auditiva, tátil, olfativa, pelo paladar, ou por uma variação e/ou combinação dos sentidos.
- Nível dois: Compreensão da situação decorrente baseado nos itens observados no nível 1, juntando seus significados conforme o objetivo, montando uma visão holística. Este nível considera a integração de muitas partes dos dados para formar a informação, de maneira a priorizar a importância e significado das informações combinadas.
- Nível três: Projeção de um futuro próximo conforme o estado dos elementos e da situação atual compreendida, este representa o nível mais alto de SAW, remetendo a capacidade de prever estados futuros.

O nível três, listado acima, só é capaz de ser alcançado através de um bom entendimento da situação, nível dois. Assim como a experiência no domínio, um sistema de informação, com interface focada em SAW, são essências para se construir uma boa consciência situacional, todavia sem a presença de um desses fatores ou falta de qualidade nos mesmos, pode levar a falhas logo nas fases iniciais, níveis um e dois, fazendo com que o humano nunca atinja o nível três.

Sendo que esses três níveis podem ser ilustrados como um núcleo, onde recebem influências externas de tarefas e sistemas, como ilustrado na Figura 2.1.

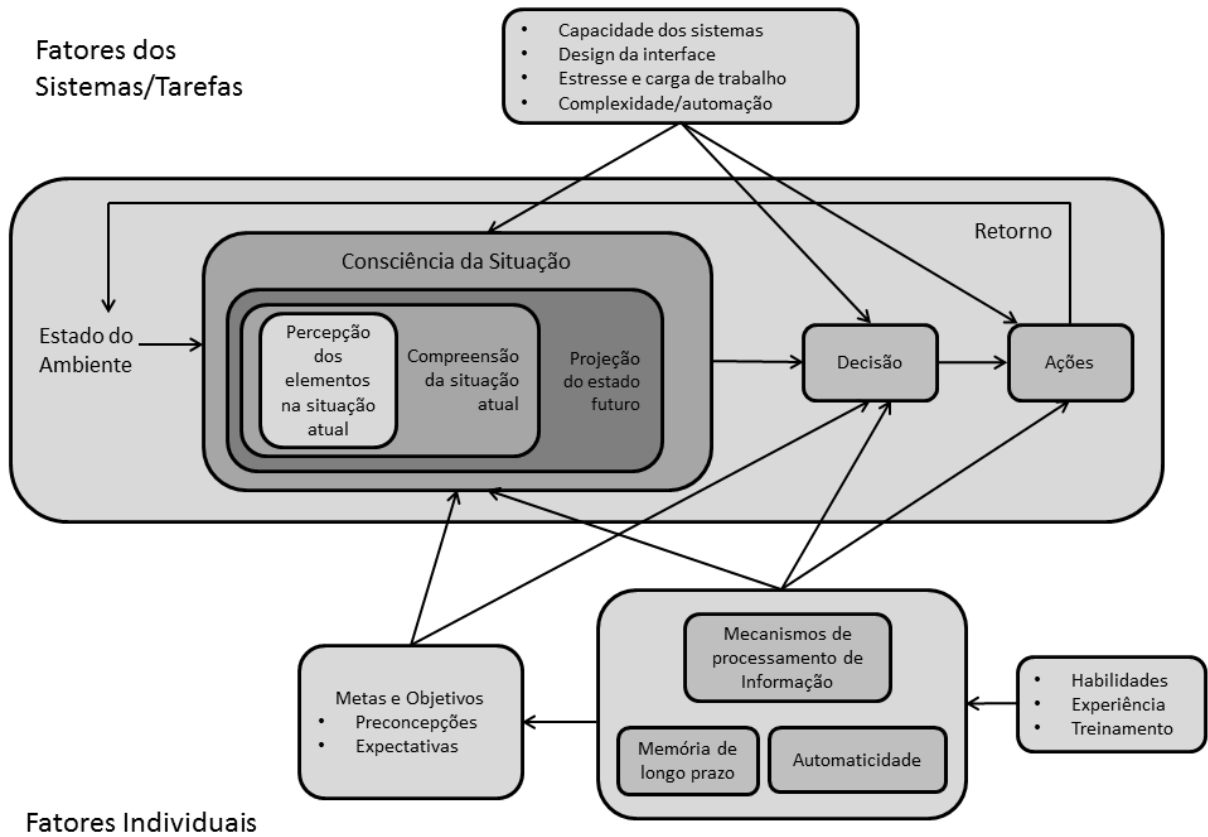


Figura 2.1: Modelo de Consciência Situacional de Endsley (Traduzido e Adaptado de Endsley, 1988)

Observando o modelo proposto por Endsley (1988), pode-se observar que o núcleo de processamento de SAW (níveis 1, 2 e 3) tem como parâmetro fatores externos adquiridos do ambiente, como sinais e estados. Também se observa um módulo de decisão externo ao núcleo de consciência, isso por SAW ser um dos principais pilares e precursores do processo de tomada de decisão.

2.2.1 Fatores Influentes Durante o Processo

Tendo que SAW é um modelo mental do operador sobre uma determinada situação em um ambiente, e baseado nas informações que ele constrói dentro desse modelo, decide o que irá fazer naquele instante. Esse modelo mental ou núcleo de SAW é diretamente alimentado por fatores externos ao indivíduo, tarefas, sistemas e ambiente físico (Kokar e Endsley, 2012).

Sendo assim, temos que tais fatores podem afetar o processo de construção e alcance de SAW do humano, facilitando ou dificultando o mesmo. Sistemas de apoio a este processo influenciam por meio de sua interface e funcionalidades, isso conforme cada tarefa executada, em torno de sua complexidade, nível de atenção, carga de trabalho e estresse provocados no indivíduo. Ainda existem outros fatores que afetam de maneira significativa este processo, como ambiente físico, com ruídos, luminosidade, temperatura, entre outros (Endsley, 2001) (Stanton *et al*, 2001).

Assim como o próprio indivíduo pode facilitar ou dificultar o seu processo de construção de SAW, pois este processo está diretamente relacionado a sua capacidade e habilidade em executar tarefas relacionadas ao domínio ou operação dos sistemas, nível de experiência, objetivos e metas. Que acabam por guiar a atenção do mesmo para pontos específicos no ambiente, sendo que cada indivíduo realiza esse processo de uma forma distinta (Kokar e Endsley, 2012) (Botega, 2015).

Para isso, sistemas e mecanismos de processamento de informação tentam se enquadrar e suprir a forma como humanos, buscam e processam informações para SAW. Através da construção de sistemas que estimulam o indivíduo a manter a atenção distribuída diante de múltiplas fontes de informação, priorizando algumas informações ou detalhes mais relevantes, que apresentam maior impacto no núcleo de processamento de SAW.

Mas ainda existem alguns fatores limitantes intrínsecos à SAW que podem interferir negativamente o processo de SAW, desde a aquisição até a evolução. Pode-se encontrar oito fatores relacionados a essas limitações, sendo eles (Endsley, 2011):

- Gargalo de atenção: Representa a incapacidade do humano em prestar atenção em mais uma informação em um dado instante.
- Limitação de memória: Humano com memória de curto prazo limitada, empregada no processo que forma a figura mental da situação atual com base na experiência e conhecimento do usuário.
- Fatores de estresse: Carga de trabalho, ansiedade e exaustão sobrecarregam e prejudicam a cognição.
- Excesso de dados: Mudança de dados rápida de dados dinâmicos dentro dos sistemas, dificultando o processo de interpretação dos humanos para manterem-se atualizados.

- Sugestões mal inseridas: Má disposição ou destaque das informações, guiando o operador para uma SAW imprecisa e fora do contexto real.
- Complexidade do sistema: Funcionalidades em excesso e maldispostas dentro dos sistemas, criam um obstáculo para a evolução do modelo mental sobre a utilização deste.
- Modelos mentais incorretos: Modelos mentais criados pelo operador que são fortemente embasados em situações familiares passados do mesmo, porém não estão condizentes com o cenário atual.
- Automação: Dependência extrema que humano pode desenvolver sobre a utilização do sistema.

2.3 Aplicação e Domínios Críticos de Consciência Situacional

SAW é um processo cognitivo realizado por todos humanos quando necessitamos compreender um ambiente ou situação para realizar uma tomada de decisão. Este fato amplia o domínio de pesquisa e estudos de SAW para quase todos os domínios, qualquer que envolva humanos, informações e a necessidade de tomada de decisão baseada nas informações.

Por isso, pesquisas que abordam essa temática são facilmente encontradas em grandes áreas críticas como a medicina, militar ou comando e controle, navegação aérea, marítima e terrestre, manutenção de equipamentos, entre outras (Endsley e Jones, 2012). Em tais domínios, a falta ou erro de SAW pode provocar a perda de vidas humanas e danos irreversíveis a patrimônios.

Trabalhos como de Gaba e Howard (1995), que tratam SAW no domínio da aviação e anestesiologia, confirmam todo o potencial de abstração e aplicação de SAW em diversos domínios. Sendo que ambos, aviação e anestesiologia, apresentam dinamicidade, complexidade, um fluxo contínuo e intenso de informações e uma taxa de risco. E também relacionam quais aspectos de uma situação o humano deve manter seu foco e atenção, sendo três aspectos: sinais sutis, situações em evolução e elementos de conhecimento especial.

Outra área que demonstra o quão importante e crítica é a SAW é a medicina, Shelton *et al* (2012) relaciona essa diretamente a segurança e vida do paciente,

onde médicos e enfermeiros devem ter uma visão holística e atualizada do status do paciente a todo instante, através de acompanhamento dos sensores e exames. Compreender essas informações e associar o seu estado atual com uma evolução é essencial para executar o procedimento mais recomendado.

Em um estudo realizado por Botega (2015), demonstra que SAW também é largamente aplicado ao domínio militar, devido à necessidade de tomada de decisões precisas em ambientes altamente dinâmicos e críticos, principalmente em relação ao tempo. De maneira que a qualidade da ação final é afetada pelo nível de consciência de um único indivíduo ou equipe sobre uma situação. Neste trabalho, os três níveis de SAW são enquadrados dentro de alguns casos, visando exemplificar o papel do humano e fatores que afetam o processo, sendo:

- Nível 1: operador ciente sobre e condição física de homens em campo, disponibilidade de recursos ou formação de tropas inimigas.
- Nível 2: compreensão do operador sobre o status das tropas inimigas, tempo até um possível confronto ou a opção de deslocamento da equipe de campo.
- Nível 3: analisa se o inimigo, na posição atual ou se continuar com direção e velocidade atuais, representa alguma ameaça, perigo ou risco à equipe, agora ou um futuro próximo.

Operadores de sistemas voltados para o atendimento de riscos, como o serviço de riscos 190 da PMESP, devem ter a capacidade de perceber algumas entidades básicas para o cenário e seus estados. No caso do serviço 190, por exemplo, pode-se ter vítimas e criminosos, suas condições físicas, descrições, atitudes e comportamentos. Considerando também as relações entre as entidades e o ambiente inserido, ainda sobre o serviço citado, pode-se caracterizar um crime, sabendo de uma pessoa está ameaçando ou ferindo outra de alguma forma, e por último, como a situação pode decorrer, novos ataques ou fuga (Botega, 2015).

Outros sistemas críticos que fazem proveito de SAW, porém mudando o modo de operação, são Sistemas de Avaliação Gestão de Risco, que buscam auxiliar especialistas, consultores e analistas a compreenderem situações e fazerem previsões em torno de um determinado ambiente, normalmente urbano, como indústrias, empresa privadas, espaços públicos e residências, possibilitando uma análise do nível de perigo e risco nestes ambientes.

Por exemplo, para realizar uma análise de risco em torno de um tipo de ambiente onde se deseja instalar uma empresa de um determinado ramo, para um especialista realizar esta análise sobre este ambiente, é necessário que ele tenha conhecimento de todo histórico de ações criminosas ou que ofereceram algum risco naquele local, com o mesmo, ou parecido, tipo de estabelecimento. Sendo que esse histórico é composto de diversos tipos de dados, como históricos policiais, imagens de câmeras, registros de redes de comunicação, entre outros.

Este processo de análise do histórico de eventos de um local para determinar o nível de risco oferecido, é também uma ação de extrema dificuldade e que exige um nível de SAW muito alto do analista, onde precisa compreender diversos tipos de informação, como estas se relacionam e neste caso, ainda considerar um fator temporal, que é o quando ação aconteceu e como esse fator reflete no momento atual que ele está. Levando em consideração que são informações de múltiplas fontes heterogêneas.

Considerando que a falta de informações pode acarretar falhas graves durante o processo de SAW, que por sua vez afeta a ação de um humano, que compromete vidas e bens ao tomar uma decisão. Com foco em tais erros e na complexidade do processo de SAW, surgem Mecanismos de Apoio ao Processo de SAW, tratados na próxima Seção.

2.4 Mecanismos de Apoio à Consciência Situacional

Partindo de erros, falhas e alta complexidade do processo e modelo de construção de consciência situacional e tomada de decisão, surgiram diversas tecnologias, algoritmo e soluções diversas que auxiliassem de alguma forma todo esse processo. As principais soluções foram Sistemas de Apoio a SAW e tomada de decisão, todavia, dependendo do domínio e dados dos quais tratados, o sistema se tornava quase ineficiente, apenas trabalhando uma interface computacional, sem considerar a qualidade da informação que estava presente, afetando assim todo o processo de SAW (Endsley, 2001) (Endsley e Jones, 2012) (Gaba *et al*, 1995) (Hall *et al*, 2006)., Com os avanços na área, aplicados ao domínio e trabalhos científicos relacionados, soluções novas e mais eficiente surgiram, agora tratando como

Sistemas de Fusão de Dados (SFD). Dos quais abordam todo o processo perante as informações que serão apresentadas, desde a aquisição de sensores físicos e não físicos, passando por processos de tratativa, inferências e apresentação a um humano, o qual pode agir diretamente sobre o processo, alterando parâmetros do processo e até mesmo complementado a informação. Fazendo com que o resultado final seja muito melhor e mais condizente com a realidade (Vincen *et al*, 2009).

Alguns desses sistemas adotam modelos de fusão que se aproximam em muito do modelo de SAW proposto por Endsley, garantindo uma sinergia entre os processos e fases, facilitando o desenvolvimento destes. E afeta também o resultado que é demonstrado ao humano tomador de decisão e a decisão que esse irá tomar, trazendo uma imagem da situação muito mais clara, atualizada e precisa. Mais detalhes dos processos de Fusão de Dados e Informações, modelos, processos e Sistemas de Fusão de Dados no próximo capítulo (Souza *et al*, 2015).

2.5 Considerações Finais

Neste Capítulo foram abordados diversos temas sobre SAW, conceitos, modelos e processos desenvolvidos e aplicados, assim como fatores que influenciam durante o processo de construção de SAW e na tomada de decisão do humano no final do processo. De forma complementar foram apresentados aplicações e domínios em que a SAW está presente e se faz de extrema importância, junto com mecanismos e sistemas de apoio à desenvolvimento de SAW e tomada de decisão, demonstrando a importância e evolução em diversas áreas da vida humana.

O Capítulo 3 está dedicado à apresentação da temática de Fusão de Dados e Informações, junto a um dos principais modelos adotados nas mais diversas áreas de estudos, o Modelo JDL. Assim como a correlação deste com o Modelo de Consciência Situacional de Endsley e quais os benefícios dessa relação, principalmente em domínios críticos que necessitam de altos níveis de SAW. Também é tratado sobre a evolução dos tipos de dados e informações que são utilizados durante os processos de fusão, especialmente sensores não físicos, e

também a alteração e adaptação dos métodos de captura, inferência e representação dessa nova classe de dados.

Capítulo 3

FUSÃO DE DADOS E INFORMAÇÕES APLICADO AO DOMÍNIO DE AVALIAÇÃO DE SITUAÇÕES DE RISCO

Este capítulo tratará sobre Fusão de Dados e Informações, demonstrando conceitos, técnicas, modelos e aplicações deste processo e como pode ser aplicado de maneira a colaborar com do domínio de situações de riscos.

3.1 Fusão de Dados e Informações

Caracterizada como um processo de transformação e inferência de dados e informações, a Fusão tem o intuito de estimar e prever estados de entidades dentro de um sistema, tendo como objetivo principal maximizar o valor da informação. Isso com redução da dimensão dos dados, incorporação de valor à informação, evolução da representatividade e produção de insumo para o processo de construção do conhecimento sobre situações. Esse processo também afeta o processo de consciência situacional, estimulando de maneira mais eficaz um humano sobre uma dada situação em um ambiente (Steinberg, 1999) (Blasch *et al*, 2013).

Segundo Llinas *et al* (2004), a Fusão de Dados e Informações é um processo multinível e multifacetado de associação de dados de múltiplas fontes, com o intuito

de captar e processar informações visando um resultado mais significativo, ao invés de consumi-las de fontes individuais, sem fusão.

Considerando o cenário tecnológico atual, com o número crescente de aplicações para os mais diversos fins. As quais produzem um número gigantesco de informações, impactando consideravelmente todo o fluxo de dados e informações. Se torna necessário algumas soluções para lidar com essa quantidade e formas distintas de dados, sendo nesse ponto que a fusão encaixa como campo de estudo multidisciplinar, aplicando conceitos e técnicas de diversas disciplinas: processamento digital de sinais, estimativa estatística, teoria de controle, inteligência artificial, e métodos numéricos clássicos.

Em um primeiro instante a definição de Fusão de Dados era dada como um conjunto de métodos para lidar com dados e informações de uma única ou diversas fontes, realizando associações, correlações e combinações de múltiplos entres estes. Com intuito de obter estimativas mais precisas sobre um determinado objeto e possíveis relações entre estes, sendo o processo um ciclo contínuo de estimativas e validações, assim como a aplicabilidade de fontes externas adicionais ou alterações no processo (Hall e Jordan, 2010).

Sistemas de Fusão de Dados e Informações compartilham de objetivos em comum, independente do domínio de aplicação, sendo estes a compreensão e predição do estado físico de uma determinada entidade. Essa análise é feita a partir das propriedades desta, características, atividade, posição espacial e movimentação, todos considerando um determinado tempo, passado, presente ou futuro (Llinas *et al*, 2004). Por exemplo, se o objetivo é analisar e deduzir o estado de um objeto, é importante não considerar apenas o objeto. Pode ser de extrema relevância determinar os atributos relacionados ao ambiente que o objeto está inserido e seu contexto, assim como as formas de interação com demais fatores presentes.

Como grande parte das linhas de pesquisa aplicadas, um dos domínios que mais incentivou e subsidiou os avanços e larga implantação de novas linhas de pesquisa, fornecendo desafios reais e complexos, foi o domínio militar. Com a Fusão de Dados e Informações o principal apoio e incentivo partiu do Departamento de Defesa Norte Americano (DoD), variando as aplicações entre sistemas de rastreamento e detecção automática de alvos (ATR), identificação de tropas amigas, inimigas e neutras (IFFN) e variantes de Comando, Controle, Comunicação e

Inteligência (C3i), visando auxiliar no processo de construção de SAW de operadores diante situações de ameaça contra vida e patrimônios (Hall e Jordan, 2010).

Assim como existem diversos modelos propostos para representar SAW, existem alguns para Fusão de Dados e Informações. Entretanto, nesse trabalho será abordado o modelo JDL de Fusão de Dados, apresentado na próxima Seção, esse modelo é uma abordagem clássica, amplamente aplicado como referência para entendimento da área e conceitos correlatos. Serão apresentados seus níveis de atuação, relação com humanos durante o processo e com SAW.

3.2 Modelo JDL de Fusão de Dados

O primeiro e mais clássico modelo de fusão de dados, veio através de pesquisadores militares do Grupo de Trabalho em Fusão de Dados e Administração Conjunto de Laboratórios (*Joint Directors of Laboratories*), junto com uma terminologia própria (Steinberg, 1999). Esse modelo foi imprescindível para superar uma época onde a área de Fusão de Dados não estava muito bem embasada. Não tendo informações e definições claras em torno de suas terminologias, algoritmos, arquitetura e processos, gerando uma dificuldade histórico para aplicação desta através das tecnologias (Rogova & Nimier, 2004).

Neste cenário, diversos pesquisadores demonstraram abordagens através de modelos, arquiteturas e frameworks, como soluções para tornar utilizável múltiplas fontes de dados e garantir uma melhor compreensão das informações presentes, suavizando as falhas e limitações de sensores. Onde surgiu uma ação colaborativa entre laboratórios liderados pelo Departamento de Defesa Norte Americano (DoD), que resultou em um modelo de referência, JDL (*Joint Directors of Laboratories*), apresentado na Figura 3.1. Com a proposta de normalizar e difundir o conhecimento sobre as bases dos sistemas de Fusão de Dados (Steinberg, 1999) (Kessler *et al*, 1991) (Linas *et al*, 2004) (Hall & MacMullen, 2004).

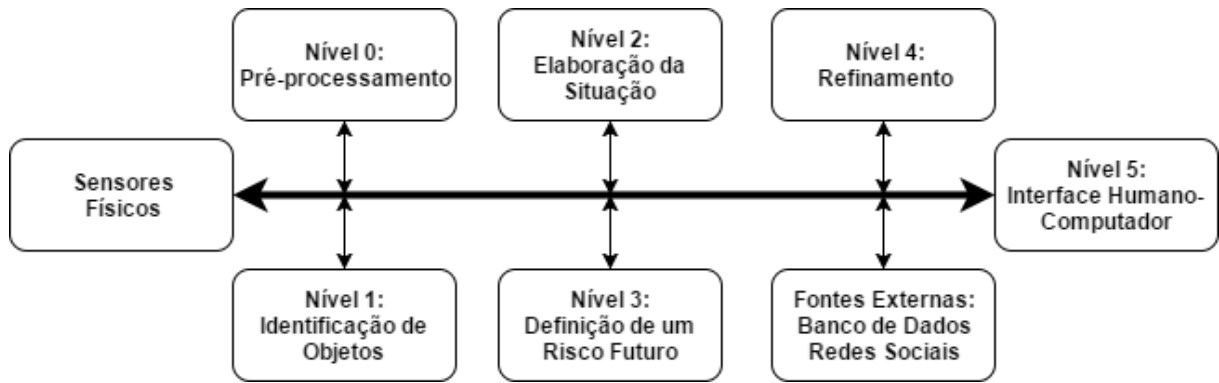


Figura 3.1: Modelo JDL de Fusão de Dados. (Traduzido e Adaptado de Hall e MacMullen, 2004)

Por ser de fácil compreensão e padronização para diversos problemas de Fusão de Dados, auxiliando também na avaliação da relevância da solução, o modelo JDL é o mais conhecido e amplamente aplicado para demonstrar a utilidade da Fusão de Dados como suporte à tomada de decisão (Hall e MacMullen, 2004).

Este modelo é composto por diferentes módulos e funções que dão suporte e possibilitam a fusão, sendo esses a preparação e processamento das informações, avaliação de objetos e situações, refinamento das informações e análise de impacto e risco. Em sua primeira e original versão, o modelo apresenta apenas quatro níveis de processos para acomodar e executar os módulos e funções citados acima, sendo eles: Identificação e Refinamento de Objetos (Nível 1), Elaboração e Refinamento da Situação (Nível 2), Definição e Refinamento de Risco (Nível 3) e Refinamento do Processo (Nível 4) (White Jr, 1991).

Todavia, o modelo não segue um processo fixo ou com determinada sequência lógica, sendo assim um modelo funcional, onde suas funções e módulos são independentes e podem alterar entre suas sequências de operação. Sendo amplamente reinterpretado e impulsionado pelo avanço tecnológico, assim como por necessidades particulares de diversos domínios onde foi aplicado (Steinberg, 2001) (Llinas *et al*, 2004) (Foo & Ng, 2013).

3.2.1 A Evolução do Modelo JDL

Mesmo sendo um modelo funcional, com uma certa liberdade entre a sequência dos processos e como estes seriam executados, houve a necessidade de uma evolução, visando melhorar e ampliar os dados de inputs do modelo e como os dados produzidos seriam apresentados e interpretados por um humano. Esta

mudança foi proposta por Steinberg *et al* em 1999, onde foram acrescentados dois novos níveis.

Sendo um dos níveis o de pré-processamento de dados providos por sensores inteligentes, este agora representado pelo Nível 0. O segundo nível proposto ficou responsável por funções específicas relacionadas as problemáticas que envolvem o humano e como esta irá consumir a informação e também em relação à interface de usuário, sendo esse o Nível 5.

Estas alterações propostas, principalmente o Nível 5, começaram a permitir uma participação mais ativa do humano durante os processos de fusão. Permitindo que o humano, por iniciativa própria, é capaz de realizar mudanças dentro dos processos de fusão, dando um *feedback* e refinando processos, reordenando e reexecutando funções disponíveis, por exemplo uma nova avaliação ou refinamento de objeto para melhorar o dado final apresentado (Llinas *et al*, 2004) (Blasch *et al*, 2013) (Botega *et al*, 2014).

Fechando assim o modelo JDL de cinco níveis o mais atual, mais difundido e aplicado, onde estes níveis se definem em (Steinberg *et al*, 1999) (Hall *et al*, 2006) (Hall e Jordan, 2010):

- Nível zero - Pré-processamento e Refinamento de Sensores: Pode ser representado como nível de *input* do processo, que visa processar dados de sensores de maneira a prepara-los para os demais níveis de fusão seguintes. Alguns exemplos deste nível são, o processamento de sinais, processamento de imagens, condicionamento de dados, transformações de coordenadas, filtragem e alinhamento em tempo e espaço.
- Nível um - Identificação Refinamento de Objetos: Processo de análise e associação de dados de múltiplas fontes ou sensores, com o intuito de definir da maneira mais confiável possível quais as entidades envolvidas no processo, como por exemplo veículos, armas ou pessoas. Também são consideradas informações fundidas, com a possibilidade de descobrir a localização ou identidade de atividades, eventos ou qualquer outra entidade que não esteja explícita em dados isolados em um primeiro instante.
- Nível dois - Elaboração e Refinamento da Situação: Este nível tem sua execução baseado nos resultados obtidos no Nível 1 do modelo JDL, o

objetivo neste é elaborar uma interpretação contextual dos dados, ou seja, analisar como as entidades encontradas no Nível 1 estão relacionadas entre elas e com o ambiente que estão empregadas. Como exemplo, um carro em movimento, onde o motorista precisa estar atento as placas, semáforos, radares e também ao estado atual do veículo, como a velocidade, a elaboração de um contexto através desde fatores, pode indicar se o motorista deve reduzir, fazer uma curva ou se pode acelerar mais. Algumas técnicas podem ser adotadas nesta etapa, como inteligência artificial, reconhecimento de padrões, *machine learning*, entre outras.

- Nível três - Avaliação de Risco e Refinamento de Ameaças: Nesta fase busca-se projetar a situação atual em um futuro próximo, através dos resultados e situação apresentada no Nível 2, para chegar em um potencial de impacto de ameaças associadas no contexto. Podemos exemplificar dentro do domínio militar, que através da identificação de ativos e condições antecipadas, este processamento identificaria e iria aferir possíveis ameaças e cursos de ação. Simulações, previsões e modelagem são algumas das técnicas aplicáveis neste nível.
- Nível quatro - Refinamento de Processos e Gestão de Recursos: Este nível fica responsável por gerenciar e validar outros processos, tentando melhorá-los sobre algum ponto, como precisão, tempo e especificidade, por exemplo. Isso através da alteração de sensores, fontes de informação e mudança de parâmetros, até mesmo técnicas. Onde envolve funções de modelagem de sensores, redes de comunicação, medidas de performance e até otimização de recursos.
- Nível cinco - Interação Humano-Computador e Refinamento Cognitivo: Busca realizar da melhor forma possível a interação da Fusão de Dados com usuários humanos, buscando dar suporte para a tomada de decisões cognitivas e ações. Estando também preocupado com as necessidades dos usuários e atender estas, focando em itens mais relevantes para operadores de Sistemas de Fusão de Dados. Algumas das funcionalidades presentes são para o uso de displays avançados, ferramentas de busca e de orientação via agentes inteligentes de software, dentre outros. Também incluem funcionalidades tradicionais

de interface humano-computador, como representação de dados e identificação de comandos de entrada.

3.2.2 Modelo JDL e o Modelo de Endsley

Com os dois modelos explícitos, é possível estabelecer uma relação entre os dois modelos, essa correspondência está entre os níveis de fusão do Modelo JDL e o Modelo de SAW proposto por Endsley. Podemos analisar essa correspondência entre os níveis 1, 2 e 3 de ambos os modelos, onde o Modelo JDL produz as entidades e demais informações fundamentais para que o Modelo de SAW tenha suas mesmas etapas cumpridas da melhor forma possível. Ou seja, os produtos gerados pelos níveis de JDL são parte dos requisitos que o Modelo de Endsley requer para causar o melhor resultado nos humanos. Assim podemos deduzir que o Modelo de SAW supri os requisitos para incluir o elemento humano no Modelo JDL, presentes no Nível 5 (Jones *et al*, 2009).

3.3 Fusão de Dados e Informações e a Participação Humana

O início dos estudos sobre processos de Fusão de Dados e Informações foram impulsionados pelo domínio militar, com o intuito de analisar e monitorar fatores militares, veículos, tropas, sistemas de armamento, entre outras. Mas não apenas coletar e apresentar esses dados, mas adicionalmente tratar esses dados de maneira a reduzir grandes volumes desses dados, deixando apenas os mais relevantes e realizando medições melhores, tendo como resultado informações mais significativas.

As primeiras soluções propostas, baseavam-se em processos e modelos voltados para o emprego exclusivo de sensores físicos e inteligência proveniente de sinais (SIGINT), algumas destas fontes são: sensores físicos dos mais variados tipos, radares, entre outros. Restringindo assim a atuação dos modelos, processos e Sistemas de Fusão de Dados à mineração e associação de dados resultantes da coleta de diversos dados destes sensores.

Esse modo de fusão é tido como Fusão de Informação de Baixo-Nível (*Low-Lever Information Fusion - LLIF*), que como dito se limita a estimar e associar dados de baixo nível provindos de sensores. Um fator de extrema importância, mas que é desconsiderado ou tratado com menor relevância nos Sistemas de Fusão de Dados que adotam essa classe de processo, é a participação de um humano durante o processo, que vai contra um dos principais objetivos que é ser insumo para o processo de construção de SAW e futura tomada de decisão (Salerno, 2002).

Esse é uma característica que está mudando com os avanços dos estudos de aplicação de Fusão nos mais diversos domínios, com demandas diferentes e também estudos para melhorar a interação humano-computador. Parte desses avanços se deve pela crescente demanda de utilização de sensores não físicos, como: banco de dados diversos que detêm conhecimento de um domínio específico, informações provindas de redes de comunicação mundial (*World Wide Web*) e principal o uso da inteligência humana (HUMINT) como input do processo e método de análise de um cenário. A adoção dessas novas fontes de informações dentro dos processos do Modelo JDL e SAW, garantem um resultado melhor para o humano, promovendo uma visão holística de um cenário (Endsley, 2001) (Blasch e Valin, 2012).

A adoção de tais sensores não físicos, apresenta grandes avanços para Sistemas de Fusão de Dados, mas também grandes desafios, pois a informação HUMINT está propícia a erros, tendências, imprecisão e omissões. Logo para ser possível adotar técnicas de fusão baseadas nessas informações, se torna necessário um nível mais alto de interpretação da informação, requer uma dedução da semântica em torno daquela informação. Onde, na maioria dos casos, não poder ser inferido e manipulado por funções automatizadas, necessitando do humano de forma ativa para interpretação e uso dessas informações. Um exemplo claro e comum desses desafios, são denúncias feitas por testemunhas ou vítimas à serviços de risco, como o 190 (Salerno, 2002) (Stampouli *et al*, 2009).

A partir deste momento, é possível observar uma alteração no papel do humano perante Sistemas de Fusão de Dados quanto a passividade no Modelo de SAW. Onde esta passa a exercer ações importantes durante o processo. Chegando em um ponto relevante, discutido por Nilsson (2012), que seria a necessidade e oportunidade do trabalho colaborativo entre humano e processos de Fusão de Informações. Junto à essas mudanças de perspectivas, pesquisadores e intuições

imersas no universo da fusão, começaram a levantar questionamentos sobre Fusão de Informações em Alto-Nível (*High-Level Information Fusion - HLIF*), a real participação do humano durante os processos, tratativas e processos as quais as entidades, provindas de sensores físicos e não físicos, são submetidas, com o intuito de adaptar e suportar as novas formas de inferência, classificação e mineração mais alto nível (Salerno, 2002) (Carvalho *et al*, 2013) (Blasch e Valin, 2012).

Botega (2015), demonstra e propõe um novo modelo de fusão de informações, onde já considera o papel ativo de humanos durante o processo de fusão e tomada de decisão, com foco em sistemas críticos para análise de situações de risco, sendo o Modelo Quantify (*Quality-aware Human-driven Information Fusion Model*). Sendo todo o modelo pensado para auxiliar o humano na construção de SAW durante a operação de sistemas críticos analisando cenários complexos, como sistemas de análise e gerenciamento de risco. Para isso, existe o emprego de diversas tecnologias, modelos e métodos para satisfazer as necessidades do humano especialista como consumidor do produto final, SAW, e também autor, fazendo um trabalho colaborativo síncrono entre homem e computador.

3.4 Adequação dos Métodos de Fusão para Suportar Informações em Alto-Nível

Costa *et al* (2012), faz uma distinção clara e direta entre processos LLIF e HLIF de fusão, onde LLIF infere e combina informações provindas de sensores para identificar, classificar e localizar objetos isoladamente. Enquanto HLIF lida com informações relacionadas a diversos objetos, considerando também o contexto que estão inseridos, com o objetivo de definir situações complexas, realizar medições sobre possíveis intenções de atores.

Um fato que se tem claro é que seres humanos tem uma capacidade muito mais avançada do que os computadores, em analisar situações ou cenários e relacionar estas, ou partes destas, em uma imagem da situação. Mas ainda apresenta um grande problema para todos os domínios, altos índices de erros, falhas e imprecisões (Costa *et al*, 2012).

Pode-se exemplificar esse fato através de um cenário de uma situação de risco, onde precisa-se levantar suposições e culpados de um sequestrado que aconteceu a três dias. Um investigador consegue observar arquivos de vídeo e imagem, e achar de maneira rápida e ágil algum relacionamento entre esses artefatos e traçar alguma relação prévia entre entidades ativas no vídeo e que podem estar nas imagens também, mas em estados diferentes. Como um homem de boné vermelho passando em frente a um banco e no momento seguinte saindo com o carro em alta velocidade. Baseado em seus conhecimentos pessoais e profissionais acumulados, ele chega a essa conclusão de maneira discreta e implícita, que pode ser o início da solução de um crime.

Para um computador processar as mesmas informações e inferir as mesmas condições, este necessita possuir o mesmo tipo de conhecimento. Mas para tornar esse todo esse conhecimento passível de entendimento de um computador, é necessário um conjunto de regras e formalismos, para que assim posso fazer o uso da melhor forma possível sempre que necessário (Costa *et al*, 2012) (Kim *et al*, 2013).

A principal e mais atual tecnologia para formalizar e explicitar conhecimentos sobre um determinado domínio, são as Ontologias. Gruber (1993), define estas como especificações da conceitualização. Auxiliando na organização, propagação, acesso, entre outros. Exemplo na Figura 3.2, demonstra uma ontologia inicial de pessoas para o domínio de gerência de riscos. Mas a principal contribuição das ontologias para o uso computacional é poder fazer com que máquinas tenham acesso e consigam organizar conhecimento e informações de uma maneira mais formalizada, mais próxima de como o conhecimento humano é estruturado, e até mesmo gerar um conhecimento melhor, em parceria com outras técnicas, como inteligência artificial e *machine learning*. McGuinness (2001), Blasch *et al* (2012), são alguns dos grandes autores que adotam ontologias em alguns dos seus trabalhos para tornar possível a realização de processos para Fusão de Informações de Alto Nível.

Na Figura 3.2, pode-se notar dois dos principais recursos presentes em uma ontologia, classes (círculos em azul escuro), e propriedades (setas com termos), onde as classes representam entidades e ativos presentes em situações e as propriedades representam como estas classes podem estar relacionadas, como uma age ou é afetada por outra. Onde a representação dessas classes e propriedades dentro de uma situação específica, representam o real contexto da ação e todos os

fatores envolvidos. Algoritmos de Fusão de Informações de Alto-Nível que adotam ontologias para inferir e representar conhecimentos, realizam estas ações em cima da presença e organização destes recursos, conforme disponibilidade. Considerando que as ontologias representam uma visão holística da situação, onde tentam prever todos os cenários possíveis dentro de um ambiente (Noy & McGuinness, 2001).

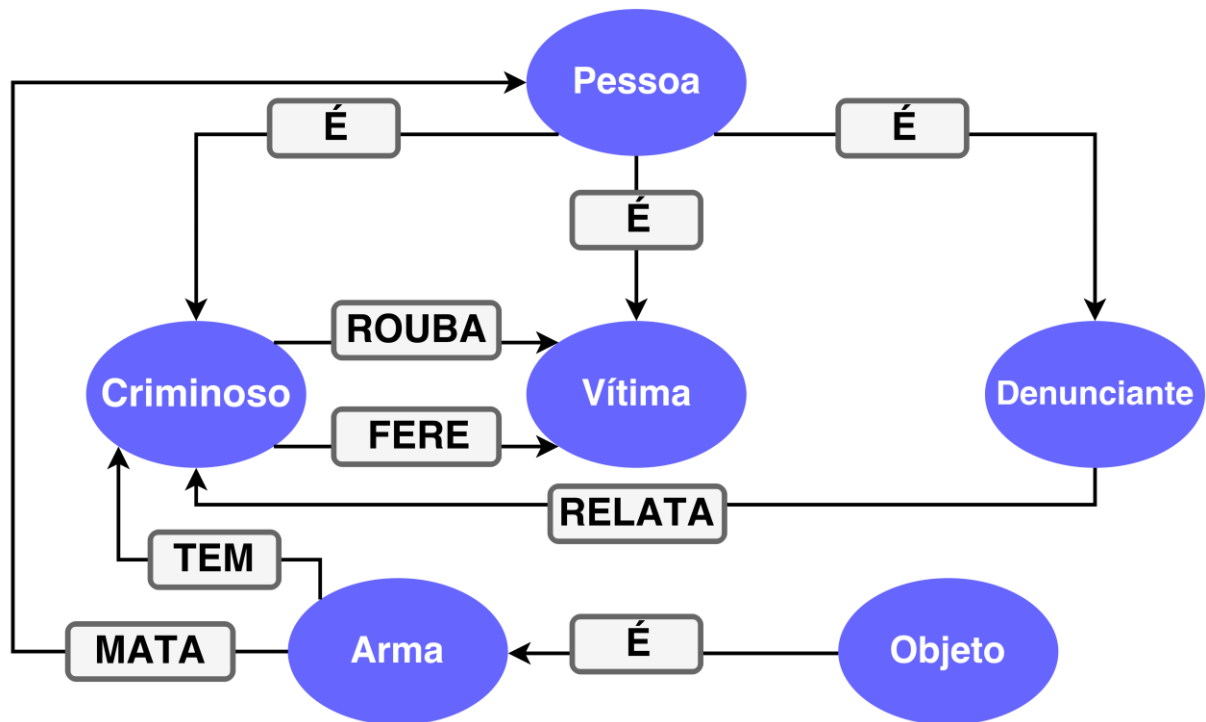


Figura 3.2: Exemplo de Ontologia

Uma das principais linguagens adotadas para representar e expressar os conhecimentos presentes em uma ontologia é a *Web Ontology Language (OWL)*, que já faz parte de outra grande área de pesquisa, a *Web Semântica*, que junto as ontologias tornam possível representar incertezas de uma informação. Outras duas tecnologias que em conjunto OWL, fazem possível a implementação de soluções baseadas na *Web Semântica*, são o *RDF* e *SPARQL*. Onde *RDF* é uma forma de descrição de entidades e vocabulários, instanciando classes e propriedades de uma determinada ontologia, possibilitando utilizar propriedades e conceitos de uma ontologia dentro de uma aplicação em tempo de execução, tendo o *SPARQL* como uma linguagem de consulta e manipulação de *RDFs* instanciados na aplicação e presentes em outras fontes de dados externas (Baader, 2005) (Carvalho *et al*, 2013) (Kim *et al*, 2013).

O emprego dessas técnicas em conjunto, são capazes de dar o poder necessários para Sistemas de Fusão de Dados de elaborar ambientes e situações mais complexos e realizar as devidas inferências antes de expor o resultado ao humano, sendo também mais flexível as possíveis interações humanas dentro do processo de fusão.

3.5 Considerações Finais

Durante este capítulo foram abordados os principais assuntos sobre Fusão de Dados e Informações relevantes para este trabalho, como conceitos base e o principal modelo adotado, JDL, o funcionamento e processos internos deste. Também foi demonstrado a evolução deste modelo, se tornando ainda mais próximo da arquitetura do modelo de Consciência Situacional de Endsley. Além dessa evolução é demonstrado o quanto ambos os modelos se aproximam e como essa aproximação torna a adoção destes em Sistemas de Fusão de Dados e Suporte a Tomada de decisão, principalmente em situações críticas.

De maneira complementar são apresentadas novas tecnologias que possibilitam o desenvolvimento de Sistemas de Fusão de Dados em mais alto nível, considerando informações provindas de sensores não físicos, como redes de comunicação e os próprios humanos.

Capítulo 4

ESTADO DA ARTE

Atualmente, existe um grande interesse no uso de sensores não-físicos para promover uma visão holística do mundo. Entre estes sensores está o uso da inteligência humana para a observação de um cenário (HUMINT), que incluem seu conhecimento como fontes qualitativas de dados e informações, direta ou indiretamente. As informações HUMINT estão sujeitas a tendências, imprecisões e omissões. Interpretar e fundir estas informações requer a dedução do significado semântico, que muitas vezes não pode ser inferido por funções automatizadas e requer a participação humana na interpretação e uso da informação. Exemplos desses desafios incluem a interpretação de relatos feitos por testemunhas (Vincen *et al*, 2009).

Recentemente, surgiram as abordagens semânticas para a fusão de dados, com o uso de ontologias como parte do processo de representação e inferência. No trabalho de McGuinness (2001), são apresentadas as primeiras classificações e organizações utilizando ontologias como parte das aplicações (ontologia de domínio), bem como o uso de ontologias externas (ontologia de topo), que provêm serviços de vocabulários e glossários controlados, contendo informações de diferentes classes, com restrições sobre os tipos e valores.

Além da evolução do papel humano durante os processos do modelo de fusão de dados, conforme relatado por Blasch *et al*. (2013). Questões referentes à qualidade dos dados e informações, também têm sido consideradas para integrar soluções de avaliação de situações de risco, em conjunto com abordagens de fusão de dados. Visando compor subsídios para melhores visualizações, interfaces de usuários e critérios de integração de dados e informações (Souza *et al*, 2015) (Botega *et al*, 2016).

Ao mesmo tempo, houve uma evolução dos modelos e algoritmos de fusão, além de novas áreas de aplicação que começaram a fazer uso de tais abordagens. Especialmente no contexto de cenários críticos e que necessitam de informações extremamente confiáveis, como as aplicações relacionadas ao domínio de gerenciamento de riscos (Vincen *et al*, 2009) (Blasch *et al*, 2013).

Carvalho *et al* (2013), faz uso de ontologias para representar áreas críticas caracterizadas pela presença de incerteza. Para este fim, foi realizado um estudo sobre técnicas capazes de representar a semântica complexa presente nessas áreas, juntamente com a incerteza que estas contêm. Como uma solução, foi proposto o uso de uma ontologia probabilística (PR-OWL) e de uma Rede Bayesiana, capazes de representar relações entre múltiplas entidades presentes em uma ontologia e ainda a probabilidade e incerteza da informação.

Costa *et al* (2013), realizou um trabalho em um domínio crítico utilizando fusão de dados sob modelos bayesianos altamente expressivos, para que seja possível processar informações de alto-nível em ambiente de fusão de dados (*High-Level Information Fusion – HLIF*). Além disso, são comparados os modelos de captura e classificação dos dados, relacionando os pontos positivos e negativos entre eles e como fatores de incerteza podem surgir e afetar os dados, considerando o papel do humano e como ele irá interagir e reagir aos dados. A partir disso, técnicas e métodos que têm a capacidade de interpretar e representar esta incerteza dos dados, dispostos em uma ontologia, foram analisados quanto às relações probabilísticas existentes entre as classes. Finalmente, o trabalho também emprega técnicas de redes Bayesianas Multi Entidade, uma técnica altamente expressiva e útil para uma grande quantidade e diversidade de dados para realizar o processo de fusão.

Kim *et al* (2013), realizou uma análise dos métodos de captura de dados e informações heterogêneas da web, utilizando o conceito de dados linkados (*Linked Data*), o que permitiu a conexão entre dados através de algum contexto em comum, assegurando a máxima captura de extensão e recuperação de dados na web. Neste trabalho, também são citados o método RDF que permite a classificação e estruturação de um documento capturado, de modo que eles podem ser reutilizados por outras técnicas ou métodos, tal como o SPARQL, o que permitiu a execução da consulta sobre a RDF gerado, permitindo um processamento maior e manipulação de dados capturados na web. Estas técnicas permitem que o operador humano

realize questões altamente complexas e receba a melhor resposta possível no contexto de fusão de dados e informações, permitindo a consulta de sinergia entre informações linkadas.

Ao analisar o estado da arte, foi possível concluir a constante evolução dos modelos e processão de fusão de dados, junto a crescente implantação em domínios críticos. Junto a preocupação de lidar com dados HUMINT e conseguir extrair todo o potencial desse tipo de informação para realizar os processos de fusão, outra tendência que pode ser notada é a preocupação com a qualidade dos dados, tanto os dados de entrada e resultantes do processo.

Porém poucos trabalhos se preocupam com a avaliação contínua de índices de qualidade, especialmente fazendo a junção de processos sintáticos e semânticos utilizando a qualidade como parâmetro e representando a mesma nas informações. Observando essas limitações quando a elaboração de um processo de fusão, que tratasse dados HUMINT, utilizando de técnicas e semânticas junto com índices de qualidade, foi o que inspirou a execução deste trabalho, visando realizar avanços da disseminação de determinados métodos, além de mitigar falhas na informação final apresenta e que será insumo para uma tomada de decisão.

Capítulo 5

DESENVOLVIMENTO DE ABORDAGEM DE FUSÃO DE INFORMAÇÕES UTILIZANDO TÉCNICAS SEMÂNTICAS PARA A MELHORIA DA QUALIDADE DA INFORMAÇÃO DE SITUAÇÕES DE RISCO

Este apresentará de maneira detalhada a metodologia adotada durante a execução do trabalho para alcançar os resultados, conforme o levantamento de requisitos utilizando GDTA, apresentando a arquitetura/processo adotado para realizar a fusão dos dados e informações. Também serão detalhados demais módulos necessários para realização da proposta, como de aquisição, avaliação e representação da informação e como estes trabalham junto ao módulo principal.

5.1 O Domínio de Gerenciamento de Risco

A Análise de Problemas Criminais no Brasil é um dos domínios com maior necessidade de aplicar o Gerenciamento de Risco. Envolvendo Situação de Risco, SAW, Qualidade da Informação, Tomada de Decisão e processamento de dados

HUMINT. Apresentando diversos problemas relacionados à qualidade da informação, além de bases de registros criminais não estruturadas, descentralizadas e não integradas. (Pereira, 2017).

Um dos maiores problemas relacionados a qualidade neste caso é a confiabilidade que se têm relacionado a informação, pois como qualquer cenário dinâmico com atores humanos durante a execução, existe a imprecisão da propagação de uma informação. Neste cenário é fácil de imaginar uma situação que acarrete problema de confiabilidade, tomemos como exemplo uma vítima mulher que visita a cidade de São Paulo e está dirigindo seu carro particular, em determinado momento sofre uma ação violenta de criminoso e tem seu carro e demais pertences roubados.

Quando esta vítima for realizar a denúncia da ação, estará alterada e nervosa, passando uma quantidade enorme de informações fora da ordem cronológica e possivelmente de maneira exagerada, estas informações são interpretadas por um humano, oficial de polícia, via telefone ou presencial, que descreve o relato dentro de um sistema de atendimento que armazena o mesmo em uma base de registros criminais do estado ou cidade. Esses dois passos da vítima realizar o relato e o humano registrar o que foi compreendido já abre uma margem para falha de qualidade, principalmente da confiabilidade e completude das informações.

Outro problema que pode afetar a qualidade do dado durante esse processo é a qualidade do sistema eletrônico utilizado pelo especialista para registrar as ocorrências, assim como as bases de dados, são sistemas segmentados por estado. Sendo que alguns desses permitem um preenchimento de campo não padronizado ou até mesmo deixar algumas informações importantes sem preencher, como por exemplo informações de local e horário, relevantes para o processo de construção de SAW e tomada de decisão de um especialista que vá consultar esses dados no futuro (Pereira, 2017).

Considerando esse cenário e a natureza dos dados HUMINT, foram desenvolvidas técnicas de Fusão de Dados e Informações visando tratar esses dados, através de processos de busca e integração de informações de fontes heterogêneas, com o foco de gerar novas informações com uma dimensão menor, porém com uma qualidade e significância muito maiores para o humano especialista que irá consumir estas informações em um Sistema de Apoio de Tomada de

Decisão, estas técnicas são demonstradas na próxima Seção, dentro de um processo e sua relação com o domínio e problemas acima.

5.2 Levantamento de Requisitos

Com a compreensão do domínio e as possíveis aplicações das técnicas de fusão, visando auxiliar no processo de análise e melhoria da qualidade das informações. Foi empregada a metodologia GDTA, útil para revelar as informações necessárias para o processo de tomada de decisão dos operadores humano durante o uso de um sistema de avaliação de situações de risco, junto com quais as tarefas que estes devem realizar.

Esta metodologia se concentra na obtenção de uma descrição precisa dos requisitos de SAW. Não apenas focada em quais dados são necessários, mas sobre como estes devem ser combinados e integrados para dar suporte a tomada de decisão do humano (BOLSTAD *et al*,2002).

Para compreender o domínio de gerenciamento de riscos, um questionário foi elaborado e encaminhado a PMESP para que fosse respondido. Analisando as respostas obtidas foi possível definir quais informações são mais relevantes para o operador, resultando no GDTA, parte deste é apresentado na Figura 5.1, a qual demonstra um de seus objetivos, que é a identificação de problemas criminais e os parâmetros considerados.

5.3 Desenvolvimento do Processo Integrado de Fusão de Informações

Todo o processo de fusão proposto neste trabalho é baseado em cima do modelo Quantify de fusão de dados, sendo que este se divide em cinco etapas: Aquisição de Dados, Avaliação da Qualidade, Fusão de Dados com Critérios de Qualidade, Representação do Conhecimento e Interface de Usuário (Botega, 2015). O foco do processo desenvolvido se mantém sobre o módulo de Fusão, destacado

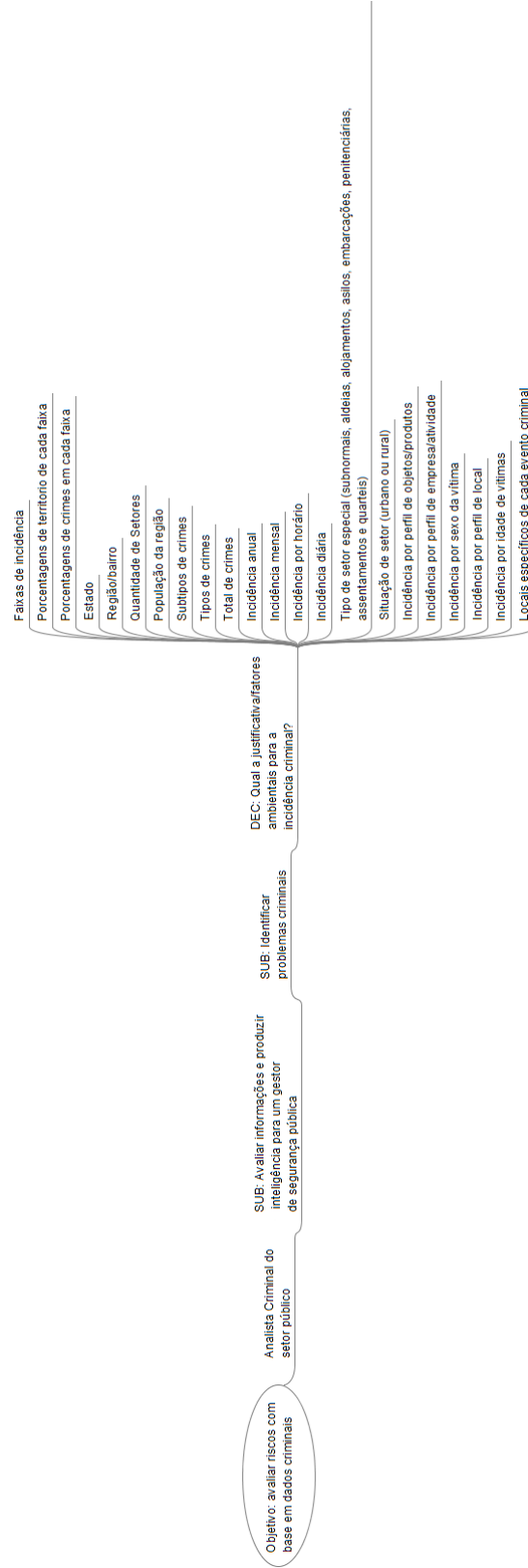


Figura 5.1: Parte da Análise de Tarefas Dirigida a Objetivos Gerado a Partir dos Resultados Obtidos no Questionário Encaminhado à PMESP

em vermelho na Figura 5.2, sendo este responsável pela parte de avaliação de objetos e situações submetidos ao modelo.

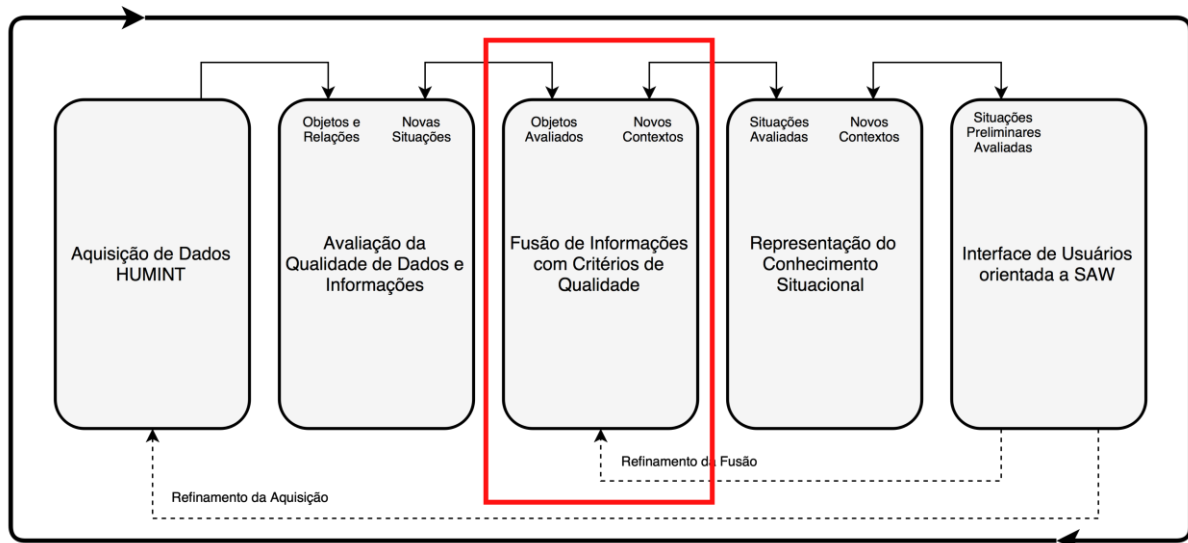


Figura 5.2: Modelo Quantify com Destaque para o Módulo de Fusão (Adaptado Botega, 2015)

5.3.1 Processo de Fusão Sintática de Informações

Considerando os desafios apresentados e as pesquisas sobre fusão de dados e suas aplicações, a primeira abordagem no domínio foi o desenvolvimento de um algoritmo de fusão considerando apenas característica sintáticas das informações para realizar os processos de busca sinérgica e integração. Essa solução analisava as informações de uma maneira mais simples, sendo que seu foco se manteve sobre os níveis 0 e 1 do modelo JDL, ou seja, específico para tratar, preparar, busca e identificação de objetos (ou entidades), considerados relevantes para a avaliação de uma situação do domínio. Identificar as principais entidades que caracterizam uma situação é importante pois serão esses que irão definir o real cenário, isoladas ou em conjunto, e serão insumo para continuidade do processo de fusão.

Esse processo de Fusão Sintática de Informações pode ser observado na Figura 5.3, detalhando cada módulo utilizado para atingir um resultado satisfatório dentro das limitações das características sintática. Esse processo será descrito em detalhes a seguir, assim como as limitações atingidas que levaram a adoção da abordagem semântica.

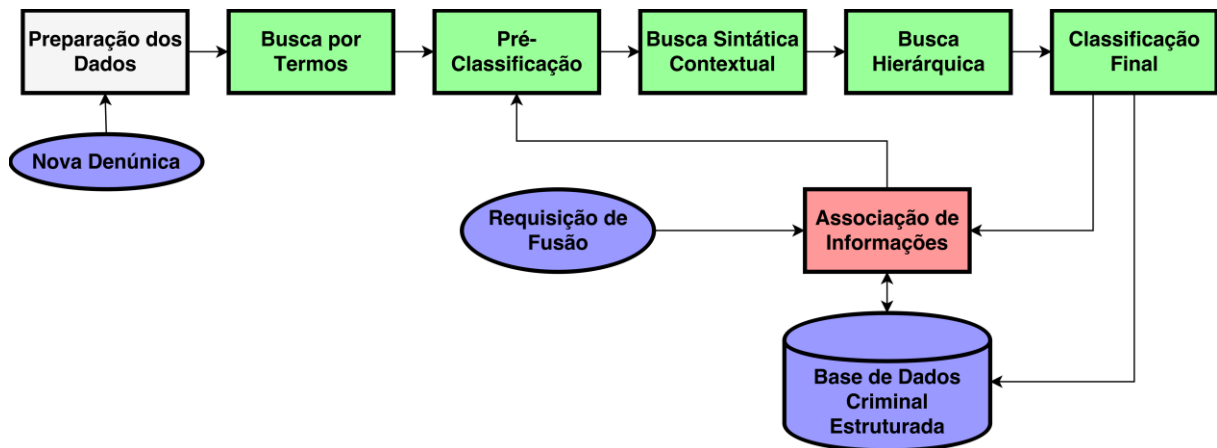


Figura 5.3: Processo de Fusão Sintática implementado para identificação e classificação de termos presentes em registros criminais

A principal entrada dessa fase, são informações HUMINT, mais especificamente denúncias presentes nas bases de dados de registros criminais. Essa entrada ao passar pelo primeiro módulo de Preparação de Dados, acontece uma limpeza, padronização e isolamento das formas gramaticais. O resultado dessa fase é uma lista de todas as palavras presentes no *input* fornecido.

Esse resultado serve como *input* para a próxima etapa, que realiza uma busca por termos relevantes que possam caracterizar um crime. Essa busca é feita de maneira sintática, ou seja, considerando a construção da palavra, isso através de consultas SQL em uma base de dados de vocabulário da com termos da língua portuguesa referentes ao cenário criminal. Essa base foi construída junto à um especialista criminal e com análise de denúncias reais, sendo que cada registro inserido considera algumas variações, como plural, gênero e até mesmo alguns erros gramáticos mais comuns.

O resultado desse processo é também uma lista de termos, porém dessa vez refinada e apenas com termos relevantes e que irão contribuir para avaliação da situação e assim com uma futura fusão que possa ser realizada, pois limita a amostragem de dados que são passíveis de covariância. Nesse ponto já pode-se notar uma grande dificuldade da abordagem sintática na problemática, caracterizado por manter um vocabulário muito amplo e ter que considerar previamente variações possíveis por um termo, afetando diretamente a eficácia de identificação de informações relevantes para uma situação.

A próxima etapa inicia o processo de análise das informações e fusão propriamente dito, a fase de Pré-Classificação, já faz uma análise combinatória dos

termos encontrados na fase anterior, isso de forma booleana, verificando a presença de determinados termos e a associação entre eles, por exemplo, a presença do termo “vítima”, “arma” e “fuga”, caracteriza um Evento Criminal de Roubo. O processo até esse estado pode ser analisado na Figura 5.4.

```
input = getDenuncia()
//padronização e isolamento gramatical
pInput = tratarInput(input)

termos = []

paraCada(pInput como termo):
    //busca o termo no vocabulário
    se (termo in db.vocabularioCriminal()):
        termos.push(termo)

paraCada(termos como termo):
    se (termo == "vítima"):
        temVítima = 1
    se (termo == "criminoso"):
        temCriminoso = 1
    ... /*para cada termo/variação no
        vocabulário é feito um teste*/

se (temVítima e temCriminoso)
    crime = 1
senão se (temFuga e temArma)
    crime = 2
... /*é realizada uma série de testes
    combinários para caracterizar a situação*/
```

Figura 5.4: Pseudocódigo demonstrando o processo inicial de Fusão Sintática

A próxima etapa vem de maneira complementar a Busca por Termos e melhorar o resultado obtido na Pré-Classificação, devido a outro desafio relacionado ao domínio e dados HUMINT. Na língua portuguesa existe uma variedade alta de combinações entre determinadas formas gramaticais e dependendo de como está feita essa combinação o sentido para situação é alterado, focado em sanar esse problema, o processo de Busca Sintática Contextual, analisa os termos em conjunto com demais termos próximos a ele, antes e depois considerando a ordem natural dos termos na denúncia.

Um exemplo, os termos “arma” e “fogo”, de maneira isolada representam informações diferentes e sem nenhuma relação, mas com a Busca Contextual, é possível terminar termos agregados, utilizando do mesmo exemplo, passaria a ser “arma de fogo”. Para isso a Busca Contextual utiliza o resultado da Busca de Termos junto com a denúncia completa apenas tratada.

Outro mecanismo que auxilia o processamento sintática é a Busca Hierárquica de termos, essa busca é realizada de maneira mais complexa ao invés de apenas identificar os termos, o objetivo é identificar termos simples ou composto presentes na denúncia, mas que não estão próximos, em sequência lógica clara ou até mesmo agindo como um gatilho para determinados casos. Por exemplo, só procurar por termos que possam identificar uma arma se encontrar um termo que represente um criminoso. Dessa forma já é possível encaixar melhor os dados o que irá facilitar o processo de avaliação final pois as informações já estarão conexas.

Na Classificação Final, as análises combinatórias são refeitas com o intuito de corrigir alguma possível falhas das informações presentes na Pré-Classificação, pois agora os dados são mais ricos e mais coesos. A classificação final dos objetos junto a todos os termos gerados e analisados, são salvos em uma base de dados estruturada e relacional, visando a estabilidade das informações e por serem altamente conexas.

A última etapa deste processo de fusão ocorre assim que a classificação é salva na base, disparando o processo de associação de informações, que representa a fusão entre as informações propriamente dita, pois nesse módulo a informação gerada durante o processo de identificação e classificação de informações (retângulos em verde na Figura 5.1) é associada e fundida, gerando uma nova situação. Essa nova situação é o resultado da fusão de uma ou mais denúncias previamente analisadas que apresentaram algum índice de covariante entre as entidades identificadas.

Com essa nova situação gerada, esta é submetida novamente ao subprocesso de classificação e identificação, porém a partir do módulo de Pré-Classificação, refinando as novas informações agora em conjunto, até que sejam novamente persistidas na base de dados. Essa parte do projeto é demonstrada na Figura 5.5, a seguir.

```

...// processo de identificação e classificação

se(temTermos e classificado):
    db.salvarSituacao(st)
    fundirInfo(st)

fundirInfo(st):
    nSituacao = []

    paraCada(st.termos como termo):
        nSituacao.push(db.buscaSituacaoPorTermo(termo))

    paraCada(nSituacao como ns):
        covarianca = 0
        paraCada(ns.termos as ntermo):
            paraCada(st.termo como termo):
                se(ntermo == termo):
                    covarianca++;
            nSituacao.covarianca = covarianca

    mcv = 0; //para analisar qual nova situacao buscada tem maior covarianca
    dSituacao = null
    paraCada(nSituacao como ns):
        if(ns.covarianca > mcv)
            mcv = ns.covarianca
            dSituacao = ns

    st.termos.push(dSituacao.termos)
    preAvaliacao(st)

```

Figura 5.5: Pseudocódigo ilustrando o núcleo da Fusão Sintática de Informações

Esse processo pode ser iniciado em duas frentes, a principal, quando uma nova denúncia é recebida que dispara o processo principal desde o início, realizando a associação apenas uma vez se conseguir identificar entidades e obter uma classificação. A outra, é sobre demanda, quando um humano ou outro sistema externo requer um novo reprocessamento das situações já presentes na base.

5.3.1.1 Estabelecendo os Limites da Fusão Sintática

Adotando apenas a abordagem sintática como processo de Fusão de Informações, é perceptível o quanto o poder do algoritmo e sua dinamicidade são

afetados, apenas relacionando os problemas pontuais citados no tópico anterior, como as inúmeras variações de um mesmo termo gramatical na língua portuguesa, em gênero, tempo e tensão, erros de escrita, vícios linguísticos, gírias e formas de construção textuais que alteram o contexto de um termo e também a ambiguidade.

Falhas ou problemas decorridos da construção das palavras, erros e variações, são passíveis de tratamento e correção pela sintática, porém se tornam muito trabalhosos a partir de um determinado momento, tornando inviável a expansão, manutenção ou otimização do processo, isso inerente ao crescimento contínuo do vocabulário.

Porém, o maior problema que a sintática enfrenta e que não consegue solucionar de maneira satisfatória, é a questão de analisar o significado de uma informação ou termo conforme o contexto que está inserido em determinado momento. Também se torna impossível determinar qual a real relação de uma informação com as demais presentes em uma situação. Para a sintática um termo sempre será analisado de uma mesma forma, apenas podendo associar este com outros termos para especificar o nível da informação.

Para exemplificar este principal limite da sintática, basta analisar a seguinte denúncia.

“acabei de ver um cara tentar assaltar um senhor aqui perto da Praça da Sé, o senhor reagiu e tirou um 38 da cintura e disparou contra o bandido, o cara fugiu pro metro, acho que tava ferido”

Realizando uma análise sintática na mesma, é possível identificar alguns termos e até mesmo obter uma classificação criminal, porém para tornar essa informação confiável computacionalmente a ponto de no futuro um especialista tomar uma decisão crítica a partir da mesma, não é possível. O resultado dos termos identificados nesta denúncia é demonstrado a seguir, os termos em vermelho são os identificados.

“acabei de ver um **cara tentar assaltar** um **senhor** aqui perto da **Praça da Sé**, o **senhor reagiu** e tirou um 38 da cintura e **disparou** contra o **bandido**, o **cara fugiu pro metro**, acho que tava **ferido**”

O processamento consegue determinar que foi um assalto, possível local e a presença de pessoa e com certeza um criminoso, mas ao pegar esses termos separadamente e tentar caracterizar uma situação, processo sintático, temos os seguintes termos: “cara”, “tentar”, “assaltar”, “senhor”, “Praça da Sé”, “senhor”,

“reagiu”, “disparou”, “cara”, “fugiu”, “metro”, “ferido”. Analisando esse resultado isoladamente, algumas questões podem ser levantadas ou desenvolvidas dúvidas no operador, como: Quem disparou? Quem ficou ferido? Quem fugiu? Os elementos encontrados não são suficientes para responder essas dúvidas para o humano, levando ele a um possível erro.

5.3.2 Desenvolvimento de uma Ontologia para Representar o Conhecimento do Domínio

A metodologia empregada para o desenvolvimento da ontologia de domínio aplicada neste trabalho, foi a metodologia 101 (ou *Ontology Development 101*) de Noy & McGuinness (2001). A mesma é constituída por sete passos, representados na Figura 5.6.

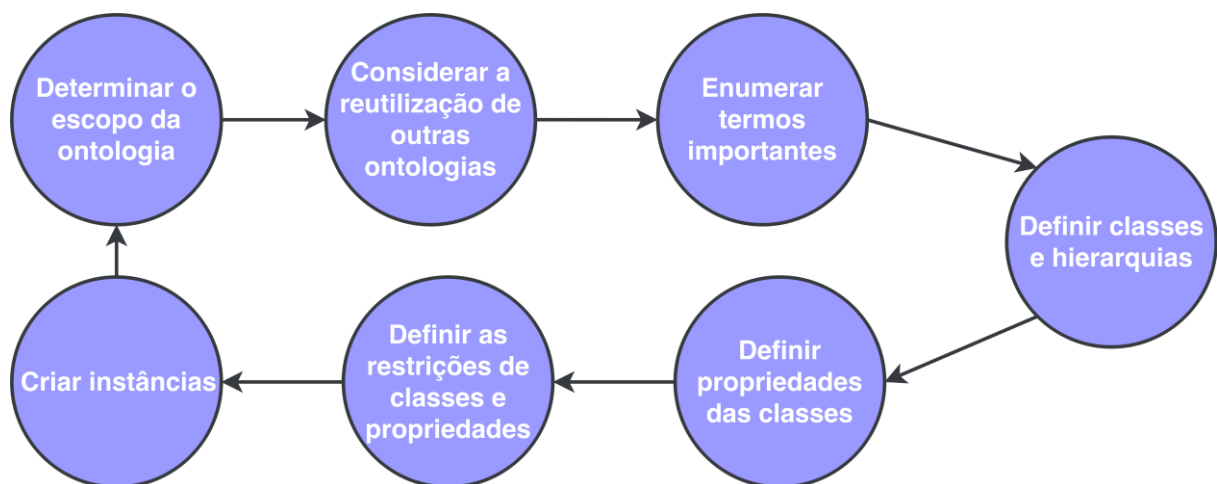


Figura 5.6: Etapas da Metodologia 101 (Traduzido e Adaptado de Noy & McGuinness, 2001)

Essa se tornou uma das metodologias mais aplicadas para construção de ontologias nas mais diversas áreas, isso devido a sua facilidade e simplicidade de ser relacionada a qualquer assunto, são passos mais de desenvolvimento abstratos que podem desempenhar seu papel em qualquer domínio.

A aplicação desta no desenvolvimento da ontologia utilizada no presente trabalho segue os mesmos sete passos, conforme descrição a seguir:

- Escopo: O domínio de aplicação e estudo dessa ontologia são situações de risco, para este trabalho situações de roubo e furto especificamente, esta será utilizada para auxiliar o processo de Fusão

de Informações que visa gerar informações mais ricas para o processo de desenvolvimento de SAW de especialistas;

- Considerar o Reuso: Nas pesquisas realizadas não foi encontrada nenhuma ontologia sobre o domínio que auxiliasse no desenvolvimento desta, principalmente considerando a utilização da Língua Portuguesa e a organização de Problemas Criminais adotada. Uma ontologia externa que foi encontrada e aplicada na construção, foi a DQV (*Data Quality Vocabulary*) utilizada para aplicar índices de qualidade dentro de uma ontologia (Referência W3C);
- Enumerar Termos: A realização desta etapa foi através da análise das respostas obtidas do questionário aplicado à operadores de sistemas críticos, além de especialistas de segurança pública e privada, junto a uma análise de denúncias presentes em bases oficiais, assim foi feito um levantamento dos principais termos encontrados e estes passaram pela validação de um especialista;
- Definir Classes e Propriedades: Baseado no questionário aplicado, foi possível identificar as informações essenciais para caracterizar uma situação de risco no domínio criminal. Com o resultado desse questionamento, foram definidas as classes principais como: Criminoso, Vítima, Local e Objeto. Dentro de cada uma dessas foram criadas classes filhas mais específicas assim como as propriedades que as caracterizam.
- Definir Restrições: Com a definição das classes e propriedades foi necessário definir as restrições do domínio, como por exemplo: “vítima pode ter um objeto”, a qual recebe a cardinalidade 1 - N dentro da ontologia;
- Criar Instâncias: A ontologia foi alimentada com diversas instancias que caracterizam as informações do domínio, estas serão apresentadas no próximo capítulo de Estudo de Caso.

O resultado final da ontologia com todas as classes pode ser visto na Figura 5.14. O desenvolvimento desta ontologia foi realizado junto ao grupo de pesquisa GIHC (Grupo Interação Humano-Computador) do Centro Universitário Eurípides de Marília.

5.3.3 Novo Processo Integrado de Fusão Sintática e Semântica

Visando solucionar os problemas e limitações da análise e inferência referentes a sintática, apresentadas no tópico anterior, foi desenvolvido um novo processo de fusão, agora composto por duas camadas de busca e associação de informações, Camada Sintática e Camada Semântica, assim como a adesão de alguns novos componentes para satisfazer as necessidades das técnicas implantadas e atingir um resultado melhor satisfatório para o processo de construção de SAW. Este processo é demonstrado na Figura 5.7, com destaque em vermelho para os módulos de fusão que foram abordados neste trabalho.

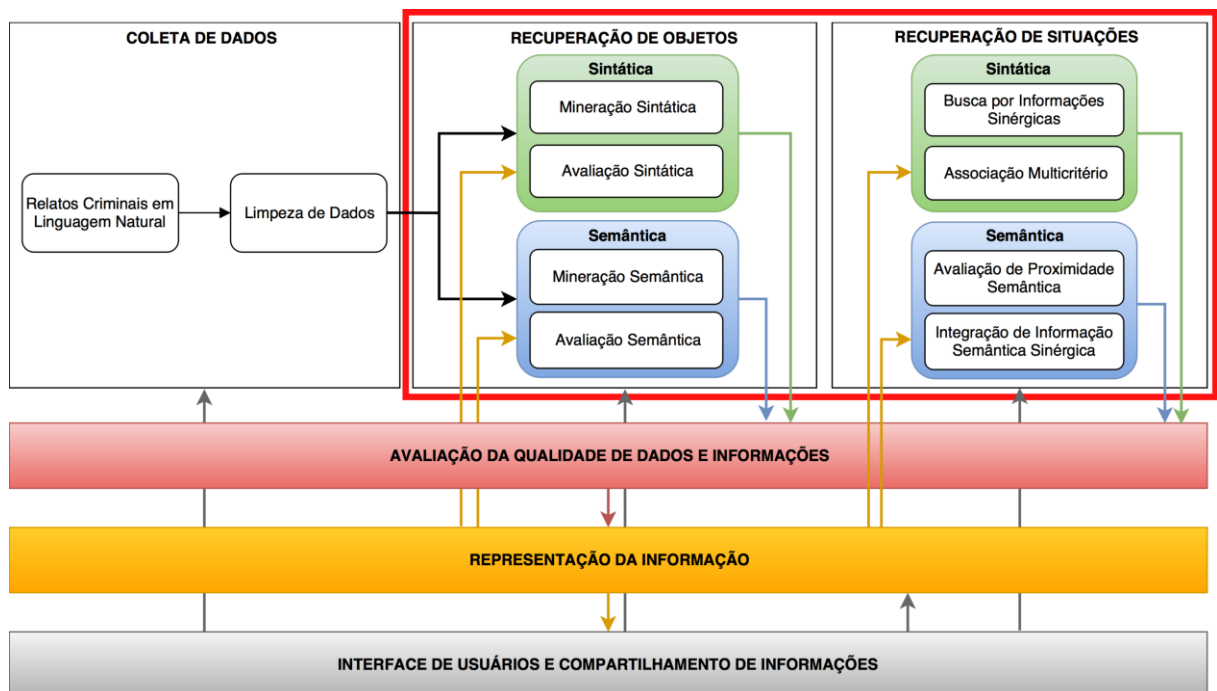


Figura 5.7: Processo Integrado de Fusão de Informações contando com Camada Sintática e Semântica (Adaptado de Junior, 2016)

Analisando o processo proposto, podemos notar a presença de diversos módulos que compõem tanto a execução sintática, quanto a semântica. Sendo que alguns destes são de uso exclusivo ou compartilhado durante o processamento, módulos de a aquisição de dados, preparação, avaliação de qualidade, representação da informação e interface de usuário são compartilhados, mas cada um com seu uso e aplicação específica, mas fica claro o núcleo de processamento

sintático e semântico. A integração de todas essas etapas, junto a ontologia desenvolvida e os métodos sintáticos e semânticos, possibilita a realização de fusão a nível de situação, ou seja, analisando a relação entre elementos identificados e o que representa, o funcionamento e integração destes módulos serão descritos a seguir. O módulo de interface de usuário não será tratado neste trabalho devido sua característica focada no tratamento e manipulação de informações.

5.3.3.1 Aquisição e Preparação dos Dados

Considerando o domínio de aplicação do projeto, existe uma variedade de fontes de aquisição de dados e informações que podem ser úteis para o processamento e tomada de decisão do operador, como fontes oficiais (abordada neste trabalho), redes sociais, sites de notícias e até mesmo canais de denúncias que fornecem áudios (Sanches, 2015).

Especificamente, para a obtenção dos objetos e atributos a serem analisados e integrados com a solução proposta neste projeto, serão empregados os resultados de um processo de aquisição de informações e processamento de linguagem natural definido por Sanches et al. (2015), o qual consiste em parte integrante do modelo de fusão de dados Quantify de Botega *et al.* (2014) (2015). Tal processo de aquisição obtém relatos HUMINT, estrutura e padroniza os dados. Esta atividade compõe o nível 1 de fusão de dados, de acordo com o modelo de referência JDL e em sinergia com o modelo Quantify.

O resultado dessas fases é um JSON (*JavaScript Object Notation*) com os termos identificados, tratados e padronizados para que possam ser consumidos pelas próximas fases, que já se dividem em sintática e semântica. Considerando o exemplo de denúncia utilizado na Seção anterior, “acabei de ver um cara tentar assaltar um senhor aqui perto da Praça da Sé, o senhor reagiu e tirou um 38 da cintura e disparou contra o bandido, o cara fugiu pro metro, acho que tava ferido”, obtemos o resultado demonstrado na Figura 5.8.

```
{
  "termos": {
    "ACABEI", "DE", "VER", "UM", "CARA", "TENTAR", "ASSALTAR", "UM", "SENHOR",
    "AQUI", "PERTO", "DA", "PRAÇA", "DA", "SE", "O", "SENHOR", "REAGIU", "E",
    "TIROU", "UM", "38", "DA", "CINTURA", "E", "DISPAROU", "CONTRA", "O", "BANDIDO",
    "O", "CARA", "FUGIU", "PRO", "METRO", "ACHO", "QUE", "TAVA", "FERIDO"
  }
}
```

Figura 5.8: Exemplo de Saída dos Módulos de Aquisição e Preparação de Dados do Novo Processo Integrado de Fusão

Este resultado da fase de Preparação dos Dados dispara dois processos em simultâneo, sintático e semântico, que darão início ao processo de busca e associação sinérgica propriamente, conforme descrito nos tópicos a seguir.

5.3.3.2 Módulo de Avaliação de Qualidade de Dados e Informações

Uma outra atividade necessária e que será executada nesta proposta é a avaliação de qualidade de dados e informações, conforme descrito em Souza *et al.* (2015). Os objetos, atributos e propriedades identificados e classificados pelas rotinas computacionais de nível 1 serão utilizadas como insumos para a busca e integração de informações sinérgicas. Entretanto, antes de serem entregues à fusão de informações propriamente dita (núcleo desta proposta), as informações serão submetidas a um serviço de avaliação de qualidade, também integrante do modelo Quantify.

Este serviço automatizado avalia informações localmente (considerando atributos e classes individuais), sob as dimensões de completude, atualidade, consistência e relevância; e globalmente (considerando relações entre objetos, denominadas situações), sob a dimensão chamada de certeza da situação, que reflete a confiança da automação na situação inferida. Ao final do processo, cada atributo, objeto e situação receberá um índice para cada dimensão considerada, que pode ser usado como parâmetro na etapa seguinte (Junior, 2016).

5.3.3.3 Um Novo Processo de Fusão Sintática Utilizando Qualidade

Com o resultado obtido nas fases preliminares é iniciado um processo de Avaliação de Objetos utilizando apenas recursos sintáticos, analisando a estrutura

de construção dos termos, considerando apenas algumas poucas variações de tensão, gênero e tempo. Este irá inicialmente minerar apenas as informações relevantes para o processo de fusão e classificação das informações, como descrição de pessoas, objetos, locais, armas e demais termos relacionados em um vocabulário de domínio desenvolvido.

Nessa fase também é executado a Busca Contextual, que relaciona termos próximos que junto representam uma informação se comparado com os mesmo de maneira isolada. Como resultado temos um novo JSON, mas agora com termos restritos as informações relevantes para o domínio e que irão influenciar no momento da fusão.

Este processo é responsável por analisar o resultado da Mineração e realizar uma classificação prévia. Que enquadra cada termo em classes de informação mais abstratas que os representam e padronizam a informação na visão sintática, que por sua vez aumenta a abrangência de reconhecimentos de padrões durante este processo. Por exemplo, ao encontrar o termo “pistola” e realizar o processo de busca e preparação esse termo é complementado com uma classe, neste caso “arma de fogo”, no processo de associação de dados sinérgicos nos próximos passos, será buscado por qualquer outro registro que tenha uma classe “arma de fogo” ao invés de especificamente o termo “pistola”.

Isto resolve parcialmente os problemas de precisão sintática, apresentados na Seção 5.3.1, pois abstrai para um nível mais alto a relação entre os dados, porém não consegue solucionar a problemática de contexto ou ambiguidade. Por exemplo, ao analisar um termo que se encaixa em duas classes, como “cano”, esse termo pode referir tanto uma arma quanto um objeto, “o bandido puxou o cano para o senhor”, “tinha um cano no meio da obra”. Só é possível definir o real sentido de “cano” analisando o contexto associado com as demais informações. Demonstrando mais um limite atingido com a análise sintática, que seria somente suprido pela aplicação de uma técnica mais poderosa, no caso um processo semântico, que será abordado na próxima Seção.

O resultado dessa fase de avaliação prévia dos objetos encontrados e que apresentem relevância, gerando uma possibilidade de organização das informações conforme as situações de risco do domínio, neste caso, roubo e furto. Sendo que essas informações são organizadas nas classes de Vítima, Criminoso, Local, Objeto e Arma, ainda utilizando de um JSON, como demonstrado na Figura 5.9.

```

{
  "termos": {
    "ACABEI", "DE", "VER", "UM", "CARA", "TENTAR", "ASSALTAR", "UM", "SENHOR", "AQUI", "PERTO", "DA", "PRACA",
    "DA", "SE", "O", "SENHOR", "REAGIU", "E", "TIROU", "UM", "38", "DA", "CINTURA", "E", "DISPAROU", "CONTRA",
    "O", "BANDIDO", "O", "CARA", "FUGIU", "PRO", "METRO", "ACHO", "QUE", "TAVA", "FERIDO" },
  "vitima": {
    "termos": { "SENHOR" } },
  "criminoso": {
    "termos": { "CARA", "BANDIDO" } },
  "local": {
    "termos": { "PRACA DA SE", "METRO" } },
  ...}

```

Figura 5.9: Resultado do Processo de Avaliação de Objetos realizado com avaliação Sintática

Este resultado é então submetido para uma das fases complementares desse novo processo que realiza a avaliação de qualidade das informações, nesse momento serão atribuídos índices locais, para cada classe identificada, e globais, para toda situação, como demonstrado na Seção 5.3.3.2. O resultado dessa situação integrada com índices de qualidade pode ser observado na Figura 5.10.

```

{...
  "qualidade": {
    "certeza": 40 },
  "vitima": {
    "termos": { "SENHOR" },
    "qualidade": {
      "completude": 15,
      ... } },
  "criminoso": {
    "termos": { "CARA", "BANDIDO" }
    "qualidade": {
      "completude": 22,
      ... } },
  "local": {
    "termos": { "PRACA DA SE", "METRO" }
    "qualidade": {
      "completude": 34,
      ... } },
  ... }

```

Figura 5.10: Exemplo de Saída com Índices de Qualidade Locais e Global

Esses índices demonstrados na Figura 5.10 são apenas simulações não representando o valor real do cálculo empregado. O resultado dessa fase,

independente do resultado da avaliação da qualidade, é persistido na camada de Representação da Informação, que é responsável, além de persistir, manter as informações em um padrão que possibilite a intercomunicação com os próximos módulos que exigem um nível mais alto de informação. São persistidos tanto os resultados da sintática quanto da semântica, porém em padrões de JSON distintos conforme a necessidade do algoritmo. A partir deste momento os relatos processados e que passaram pelos módulos anteriores são classificados como situações, até mesmo aquelas que não apresentam um nível muito alto de complexidade ou detalhamento de objetos.

O próximo passo da execução sintática é iniciar a Avaliação de Situações onde os dados serão fundidos a nível de outras situações também presentes na camada de Representação da Informações. Na abordagem sintática alguns algoritmos e critérios serão utilizados para determinar a sinergia entre as informações e associar (fundir) estas, que será descrito com mais detalhes nos próximos parágrafos.

Com a informação na camada de Representação, o próximo passo é a execução de uma Busca por Informações Sinérgicas, em situações previamente processadas, que têm algum grau de covariância em sua sintaxe, ou nas classes presentes. Para realizar essa busca por valores sinérgicos como a estrutura da palavra, são consideradas pequenas variações na estrutura, como tensão, gênero ou radical. Essas pequenas variações podem ser relevantes para impulsionar o poder desta busca. Por exemplo, se uma situação tem que "começar a correr" e outra situação anterior tem "um cara correu", esta informação será julgada como sinérgica, por causa do radical "corridas". Após esta busca sinérgica para outras situações, aqueles que tinham algum grau de sinergia são agrupados em conjunto de dados.

O algoritmo empregado internamente para a Busca Sinérgica é o KNN (*K-Nearest Neighbor*) (Castanedo, 2013), porém com algumas alterações em sua estrutura para que apresentasse o melhor resultado o domínio (Pereira, 2017), extraindo dados dos relatos em diversos níveis de abstração para a identificação de classes, subclasses e valores associados.

As regras de sinergia sintática são construídas sobre as subclasses e valores encontrados. As classes são utilizadas para verificar a existência das subclasses e validar o conjunto de regras inerentes ao domínio. Tendo como resultado uma

seleção de outras situações que apresentam algum índice de covariância, ou proximidade (termo utilizado no algoritmo de KNN). As Figuras 5.11 e 5.12 representam respectivamente, uma demonstração do algoritmo de Busca de Informações Sinérgica e o resultado obtido com o mesmo.

```

BuscaSinergica(st):
  situacoes = []
  paraCada(st.classe como cl): //a qualidade também é considerada uma classe
    situacoes.push(db.buscaSituacaoPorClasse(cl))
    //busca no banco de registros qualquer outra situação que tenha classes
    em comum
  indice = 100/st.classes.quantidade;
  paraCada(situacoes como ns): //analisa cada situacao encontrada
    covariancia = 0;
    paraCada(ns.classe com ncl):
      paraCada(st.classe como cl):
        se(cl == ncl)
          covariancia += indice
    ns.covariancia = covariancia
  retorno situacoes

```

Figura 5.11: Pseudocódigo demonstrando a base da Busca de Informações Sinérgicas

```

{
  "situacoes": {
    "0":{
      "criminoso":{...},
      "vitima":{...},
      "qualidade":{...},
      "covariancia": 40.6 },
    "1":{
      "criminoso":{...},
      "vitima":{...},
      "qualidade":{...},
      "covariancia": 10 },
    ... }
  }

```

Figura 5.12: Demonstração de Resultado de uma Busca de Informações Sinérgicas

Esses conjuntos ou coleções de informações geradas passam então para uma fase denominada Associação Multicritérios, que considera para a fusão, além

dos conjuntos de dados sinérgicos, os índices de qualidade atribuídos aos dados e informações inferidas. Assim selecionando apenas uma única situação a ser fundida com a atual, aquela que apresentou o maior índice de covariância. Esta atividade organiza e vincula à situação corrente, as novas informações de acordo com limiares de qualidade estabelecidos nos requisitos, para que o produto seja uma situação única, completa e a mais precisa possível. A Figura 5.13 contém um pseudocódigo que representa parte do algoritmo de associação.

```
Associacao(st):
    situacoes = BuscaSinergica(st)

    mcv = 0;
    //para analisar qual nova situacao buscada tem maior covariancia
    dSituacao = null

    paraCada(situacoes como ns):
        se(ns.covariancia > mcv):
            mcv = ns.covariancia
            dSituacao = ns

    paraCada(dSituacao.classe como cl):
        escolha(cl):
            caso "vitima":
                st.vitima.push(cl.termos.vitima)
            case "criminoso":
                st.criminoso.push(cl.termos.criminoso)

    retorno st
```

Figura 5.13: Pseudocódigo Demonstrando o Core da Associação Multicritério

Ao final do processo, temos como resultado uma nova situação, sendo esta enviada novamente para a camada de avaliação da qualidade de dados, para ser avaliado e obter novos indicadores de qualidade. Aquelas situações que não atingiram o nível básico de qualidade são armazenadas temporariamente para que possam ser comparados com outros resultados, que são gerados pelo processo de fusão de informação semântica, que ocorre em paralelo com este.

5.3.3.4 Processo de Fusão Semântica utilizando Qualidade

O Processo de Fusão utilizando de técnicas semânticas consome o mesmo insumo da parte sintática que foi produzido pelos módulos de Aquisição e Preparação, descritos na Seção 5.2.3.1. Todavia para viabilizar e extrair o maior poder de processamento semântico, foi construída uma ontologia para representar os conhecimentos relacionados ao domínio da melhor maneira possível, possibilitando assim que um computador analisasse as informações relacionadas à uma situação de uma maneira mais inteligente, considerando contexto e relacionamentos entre informações, ao invés de apenas a presença de alguns determinados elementos, como é feito na sintática.

Essa ontologia foi desenvolvida com base em conhecimentos sobre o domínio, mediante consulta a especialistas criminais parceiros e análise de situações reais disponibilizadas para realização deste trabalho, sendo guiada pela metodologia de desenvolvimento 101, melhor descrito na Seção 5.2.2.

Contendo diversas classes, onde estas representam os principais elementos presentes nas denúncias analisadas, respeitando questões como hierarquia e simetria as classes e suas propriedades, por exemplo, existem as classes, Criminoso, Vítima e Denunciante, todas essas classes são referentes a uma pessoa, então existe uma classe Pessoa que está diretamente ligada a classe de Situação que irá definir o papel de cada uma das classes filha de Pessoa, a ontologia completa pode ser vista na Figura 5.14, a Tabela 5.1 descreve com detalhes as classes presentes nesta.

Tabela 5.1: Tabela de Classes da Ontologia, descrevendo o que cada classe presente representa para a situação.

Descrição das Classes da Ontologia	
Classe	Descrição
Situação	Classe de nível mais alto que representa toda a situação possível de ser representada utilizando esta ontologia, sendo específica pela classe Crime
Crime	Esta classe é responsável por definir quais são as possíveis situações a serem representadas, no caso, os crimes

	de Roubo e Furto
Roubo	Classe que representa a especificação de situações de Roubo, representado quais as propriedades com as demais classes que caracterizam um roubo para o domínio
Furto	Classe que representa a especificação de situações de Furto, representado quais as propriedades com as demais classes que caracterizam um roubo para o domínio
Pessoa	Responsável por agrupar os possíveis tipos de pessoa que estão presentes nas situações de Roubo e Furto, no caso, Criminoso, Vítima e Denunciante
Criminoso	A classe Criminoso representa o elemento a Pessoa que atua nos crimes, realizando determinadas ações sobre a Vítima, na maior parte das vezes realiza a ação com algum elemento de Arma, subclasse da classe Objeto
Vítima	Representa a Pessoa que sofre com a ação representada, Roubo ou Furto, sendo influenciada pelas propriedades com criminoso, na maior parte das vezes está associada a algum objeto que sofre a ação em conjunto
Denunciante	Classe que representa o papel de alguém que apenas relatou uma situação, em alguns dos casos a classe Vítima assume esse papel em conjunto
Local	Classe responsável por representar os locais da situação, estando diretamente

	relacionada as classes de Roubo e Furto
Objeto	Demonstra qualquer objeto que esteja relacionado as ações de Roubo e Furto, tanto os do Criminoso, principalmente arma, quanto da Vítima
<i>QualityGraph, Quality Measurement Data, DataSet, Category, Concept, Observation, Metric, MeasureProperty e Dimension</i>	Classes utilizadas para representar informações dos índices de qualidade avaliados, pertencentes ao vocabulário <i>Data Quality Vocabulary (DQV)</i>
Ensino, Entretenimento e Transporte	Classes externas utilizadas para enquadrar os tipos de locais
Droga, Eletrodoméstico, Documento, Obra Arte, Veiculo, Munição, Arma, Vestuário, Acessório e Eletroeletrônico	Subclasses de Objeto utilizadas para especificar qual tipo está presente na Situação

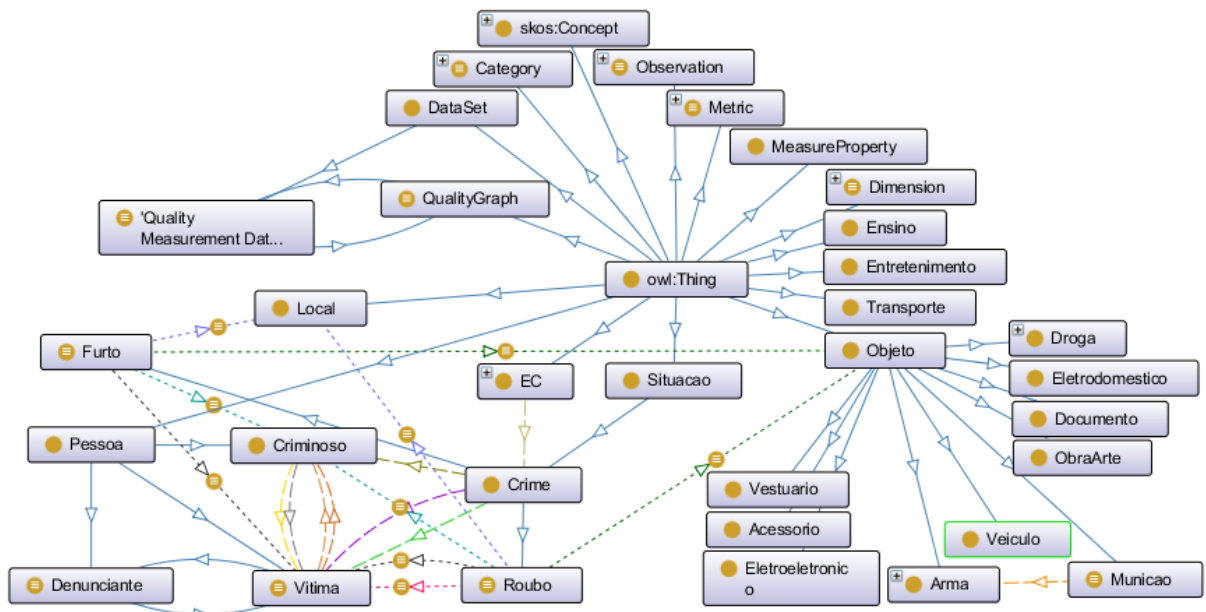


Figura 5.14: Ontologia Desenvolvida para o Domínio de Gerenciamento de Riscos (Roubo e Furto)

Com a adoção da ontologia nos processos internos de busca e associação de informações, foi necessária uma alteração nos padrões de comunicação dos módulos, o JSON, utilizado no processo sintático, não consegue representar todas

as informações que a ontologia irá descrever afetando diretamente a qualidade final da informação dentro dos módulos. Assim, o novo padrão de transmissão de informações definido para essa linha de execução foi o *JavaScript Object Notation for Linked Data* (JSON-LD), que foi construído especificamente para aplicações que aplicam soluções com dados conectados e semânticos (Sporny *et al*, 2014) (Isotani & Bittencourt, 2015).

Outra grande alteração que ocorre em relação a representação das informações, são os índices de qualidade, que passam a respeitar um vocabulário aplicado na construção da ontologia, sendo este o *Data Quality Vocabulary* (DQV). Um vocabulário desenvolvido pelo Grupo W3C (*World Wide Web Consortium*), visando estimular e facilitar a publicação, intercâmbio, consumo e expressividade de metadados de qualidade presentes em conjuntos de informações.

Esta ontologia será utilizada de duas maneiras, a primeira maneira é como uma ontologia de representação do domínio contando com diversas instâncias (referentes os termos utilizados para a classificação sintática). Onde as consultas durante o processo de fusão irão buscar o significado de uma informação, retornando as classes e propriedades relacionadas aquela a determinada instância consultada. A segunda forma de uso da ontologia, é durante a etapa de associação de informações, onde será consultada novamente. Porém agora para consultar apenas as propriedades entre as classes e tentar descobrir as reais relações entre os objetos encontrados durante as buscas sinérgica, estabelecendo assim uma única situação final mais concisa.

Com todas as alterações necessárias para suportar a abordagem semântica durante todo o processo, a primeira fase de execução pós aquisição e preparação, comum para as duas abordagens, sintática e semântica, é a avaliação de objetos, que parte do mesmo conceito da execução sintática, identificar e categorizar os principais elementos que caracterizam uma situação de risco, porém agora com técnicas semânticas.

O insumo para essa parte é o mesmo da parte sintática, conforme demonstrado na Figura 5.6, mas ao invés de realizar uma consulta SQL simples em uma base de vocabulário com diversos termos associados de maneira simples, é realizada uma consulta SPARQL nas instancias presentes na ontologia, conforme descrito acima. Essa consulta SPARQL inicial, ainda que simples, apenas para caracterizar um termo, traz resultados mais relevantes se comparada a SQL, pois já

agrega as propriedades da instancia identificada, onde as Figuras 5.15 e 5.16 demonstram respectivamente, uma consulta SPARQL para identificar as possíveis ações, ou propriedades do criminoso sobre uma vítima.

```

1 SELECT ?action
2 WHERE {
3   ?action rdfs:range vitima: .
4   ?action rdfs:domain criminoso:
5 }|

```

Figura 5.15: Exemplo de Consulta SPARQL para Inferir as Propriedades Sobre Determinadas Classes

action
voc:furta
voc:rouba

Figura 5.16: Resultado da Consulta SPARQL Realizada na Figura Anterior

Este processo de consultas SPARQLS, é compatível com o módulo de Mineração Semântica, descobrindo todas as classes e propriedades relacionados aos termos identificados. Este resultado de diversos SPARQLS é analisado no módulo de Comparação Semântica, associando as entidades entre elas mesmas buscando caracterizar uma situação da melhor maneira possível, um exemplo da estrutura desse resultado utilizando um JSON-LD pode ser observado na Figura 5.17.

```

{"@context": {
  "objeto": "http://localhost:8090/objeto#",
  "local": "http://localhost:8090/local#",
  "pessoa": "http://localhost:8090/pessoa#",
  "criminoso": "http://localhost:8090/criminoso#",
  ...},
"@id": "http://localhost:8090/situacao#st01",
"@type": "crime:roubo",
"vitima:genero": "Mulher",
"criminoso:rouba": "vitima:",
...}

```

Figura 5.17: Exemplo de Resultado da Avaliação de Objetos com Abordagem Semântica

Este JSON é então submetido para a camada de Avaliação de Qualidade, nesse momento serão atribuídos índices locais, para cada classe identificada, e globais, para toda situação, como demonstrado na Seção 5.2.3.2. Porém o resultado desse processo não será apenas referenciado em um JSON simples, como demonstrados na Figura 5.9, esta passa a respeitar a estrutura proposta pelo DQV para que esta informação possa também ser considerada como uma classe e critério de associação durante o processo semântica de Avaliação de Situações. A Figura 5.18 demonstra como fica organização dos índices de qualidade utilizando do DQV para representação.

Com a situação já definida com suas respectivas entidades e índices de qualidade, está é persistida na camada de Representação de Informações que está hábil a manter todas as propriedades semânticas das informações e suas estruturas internas, caracterizando assim uma situação. O próximo módulo executado nesse momento já é referente a fase de Avaliação de Situações, que neste caso irá avaliar a covariância, ou nível de aproximação semântica.

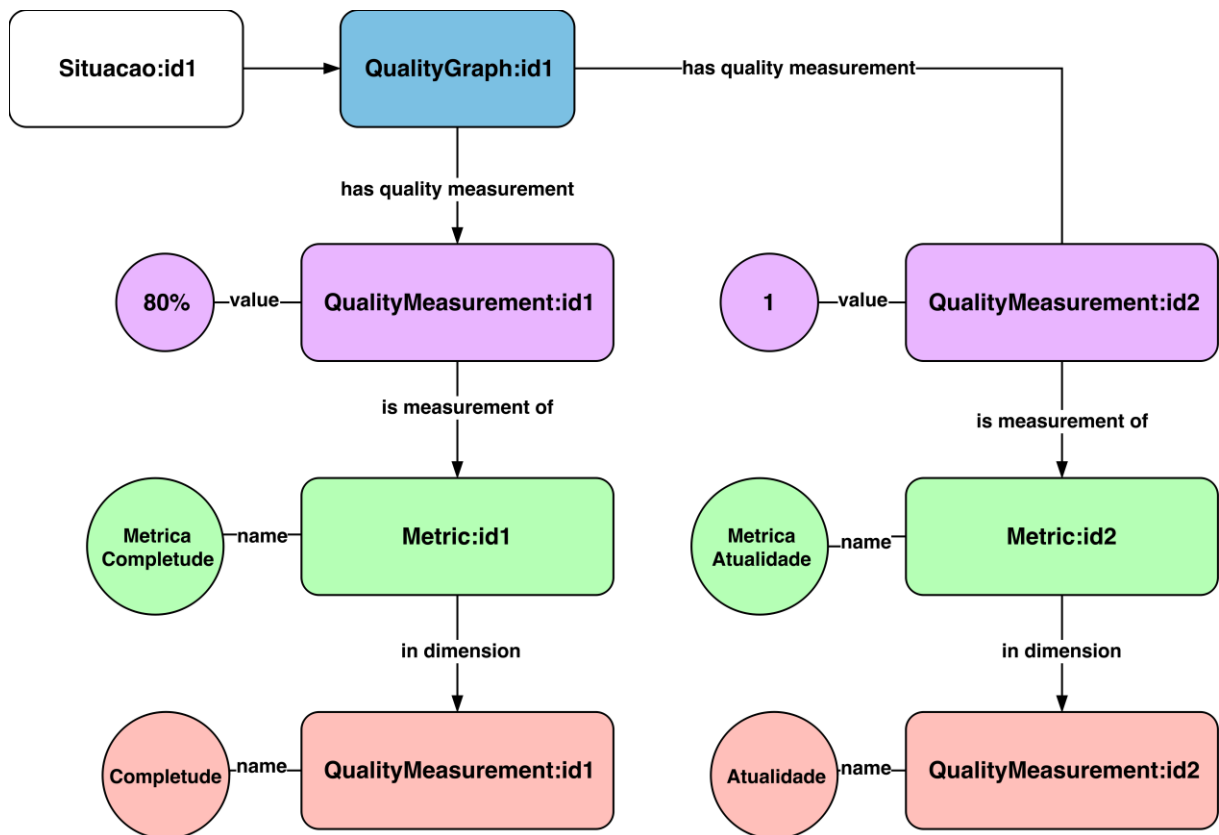


Figura 5.18: Exemplo de Representação de Índices de Qualidade Utilizando DQV

Esse módulo irá fazer uma seleção de as situações previamente processadas pelo algoritmo semântico e que estão presentes na camada de representação da informação. Para cada situação encontrada que tenha algum nível de aproximação, analisando classes ou propriedades presentes, é realizado um processo comparativo de classes para classe, propriedade para propriedade e o mais completo, classes e propriedades em determinados conjuntos, analisando padrões (contextos) buscando por informações similares. O resultado desse processo é um conjunto apenas com situações similares em algum grau com a atual, como demonstrado na Figura 5.19.

```
{
  "@context": {
    "objeto": "http://localhost:8090/objeto#",
    "local": "http://localhost:8090/local#",
    "pessoa": "http://localhost:8090/pessoa#",
    "criminoso": "http://localhost:8090/pessoa#Criminoso",
    "vitima": "http://localhost:8090/pessoa#Vitima",
    ...
  }
},
"@graph": [
  {
    "@id": "http://localhost:8090/situacao#st02",
    "@type": "crime:roubo",
    "vitima:idade": "Adolescente",
    ...
  },{
    "@id": "http://localhost:8090/situacao#st05",
    "@type": "crime:roubo",
    "criminoso:sexo": "Masculino",
    ...
  },{
    "@id": "http://localhost:8090/situacao#st10",
    "@type": "crime:roubo",
    "objeto:Documento": "CNH",
    ...
  } ]
}
```

Figura 5.19: Demonstração de Resultado de uma Busca de Informações Sinérgicas através da Semântica

Esse novo conjunto de situações é submetido ao processo de Integração de Informações Sinérgicas, onde irá definir qual das informações tem a maior aproximação com a atual, essa análise é feita com base nas classes, propriedades e contextos presentes, sendo que aquele que apresentarem a maior sinergia, serão fundidos com as informações atuais. Essa associação semântica deixa de fundir termos e classes em JSON, para associar instancias, classes e propriedades presentes, confirmando informações presentes na situação ou até mesmo criando propriedades inexistentes. O core do algoritmo de Integração Semântica pode ser visto na Figura 5.20.

```

IntegracaoSemantica(st):
    situacoes = BuscaPorProximidade(st)
    mcv = 0;
    dSituacao = null

    paraCada(situacoes como ns):
        se(ns.covariancia > mcv):
            mcv = ns.covariancia
            dSituacao = ns

    paraCada(dSituacao.classe como cl):
        se(existe cl.propriedades):
            paraCada(cl.propriedades como pr):
                AssociacaoPropriedades(pr, cl, st)
                /* montar contextos utilizando cada classes analisada
                utilizando suas propriedades em relação
                a situação atual*/
            paraCada(cl.subclasses):
                BuscaPorSubClasses(cl, st)
                // busca por classes filhas mais especificas
            paraCada(cl):
                BuscaPorClassesAssociadas(cl.propriedades, st)

    //Cada Método de Busca Chamado Realizada Consultars SPARQL

```

Figura 5.20: Pseudocódigo de Integração Semântica de Informações

Após a associação das informações e a instanciação de uma nova situação, está será novamente enviada a camada de Avaliação de Qualidade, para que os índices sejam recalculados conforme o estado atual. Com todas as informações atualizadas e devidamente estruturadas, essa é enviada para a camada de

Representação da Informação, onde será persistida até que seja associada com outras possíveis situações em execuções futuras.

5.4 Considerações Finais

Este capítulo descreve a execução para atingir a proposta deste trabalho, são demonstrados diversos processos de fusão de informações, desde o primeiro teste com fusão sintática, os limites atingidos e quais as soluções adotadas até chegar à um processo integrado novo de fusão sintática e semântica. Com isso são detalhados os módulos associados e como acontece uma execução comum de um processo de fusão completo, demonstrando os padrões de comunicação adotadas e qual o funcionamento dos principais módulos. Deixando claro a necessidade da abordagem semântica no domínio.

Capítulo 6

ESTUDO DE CASO COM O PROCESSO INTEGRADO DE FUSÃO DE INFORMAÇÕES

Neste capítulo será realizado um ensaio dentro de um ambiente controlado, porém com dados reais, utilizando o Processo Integrado de Fusão de Informação, com as técnicas sintática e semântica, provando a eficiência da ambas conforme a eficiência, porém evidenciando o quanto a semântica apresenta melhoras quando considerado o domínio de risco e dados provenientes de humanos.

Para realizar este estudo serão considerados três denúncias, todas referentes a uma mesma situação de roubo, porém com visões e informações distintas, que isoladas não são tão completas e nem oferecem uma SAW muito elevada para o operador final. As denúncias são as seguintes:

Denúncia 1: *“a vítima parou o veículo em um semáforo, foi surpreendido por dois indivíduos que estavam em uma motocicleta preta. O garupa bateu com uma arma no vidro e proferindo grave ameaça subtraiu o veículo da vítima. Os autores evadiram-se com destino ignorado. Vítima não possui condições em descrever os autores com detalhes.”*

Denúncia 2: *“dois caras numa moto de cor escura roubaram um carro no farol do cruzamento da Rua da Mooca com a Taquari, apontaram um revolver e levaram o carro da mulher. Um deles estava de casaco vermelho.”*

Denúncia 3: *“dois meliantes roubaram um carro de cor cinza na rua da Mooca ao lado do banco Santander. Os meliantes estavam numa moto de cor preta um deles estava de blusa azul e calça jeans partindo sentido hospital villa lobos.”*

O ciclo de execução antes da fusão sintática ou semântica, se resume em: identificação de objetos, camada de representação, avaliação de qualidade e após os dados avaliados, retornam para camada de representação. A camada de representação da informação é utilizada para auxiliar os processos de fusão durante a execução, mantendo sempre a versão mais atual de uma informação, com conteúdo e índices de qualidade condizentes.

A primeira fase de todo o processo, independente se será avaliação sintática ou semântica, é a identificação de entidades e objetos presentes nas denúncias, este processo é baseado em técnicas de processamento de linguagem natural, utilizando de um vocabulário rico da língua em questão e focado no domínio de aplicação. O resultado deste pode ser visto a seguir, onde os objetos identificados estão destacados em vermelho.

Denúncia 1: “a **vítima** parou o **veículo** em um semáforo, foi surpreendido por **dois indivíduos** que estavam em uma **motocicleta preta**. O **garupa** bateu com **uma arma** no vidro e proferindo **grave ameaça subtraiu** o **veículo** da **vítima**. Os autores evadiram-se com destino ignorado. **Vítima** não possui condições em descrever os autores com detalhes.”

Denúncia 2: “**dois caras** numa **moto** de **cor escura roubaram um carro** no farol do cruzamento da **Rua da Mooca com a Taquari**, apontaram **um revolver** e levaram o **carro** da **mulher**. Um deles estava de **casaco vermelho**.”

Denúncia 3: “**dois meliantes roubaram um carro** de **cor cinza** na **rua da Mooca** ao lado do **banco Santander**. Os **meliantes** estavam numa **moto de cor preta** um deles estava de **blusa azul** e **calça jeans** partindo sentido **hospital villa lobos**.”

Após a fase de identificação de entidades e objetos, ocorre a distinção entre os processos sintático e semântica. Neste momento as situações começam a ser caracterizadas para então serem fundidas. No sintático cada esse processo é realizado pelos módulos de Mineração e Comparação Sintática, o resultado desses é então submetido às camadas de Avaliação de Qualidade e Representação da Informação. A camada de qualidade, é chamada para uma avaliação mais simples, instanciando os primeiros índices de qualidade tanto locais, para cada elemento e um global de certeza. Na camada de representação nesse momento as informações estão mantidas em um Objeto JSON (*JavaScript Object Notation*), mesmo que não

sejam dados muito precisos ou qualificados, neste momento são caracterizados como situações, como demonstrado nas Figuras 6.1, 6.2 e 6.3.

```
{
  "termos": {
    "VITIMA", "VEICULO", "DOIS", "INDIVIDUOS", "MOTOCICLETA", "PRETA", "GARUPA",
    "UMA", "ARMA", "GRAVE", "AMEACA", "SUBSTRAIU", "VEICULO", "VITIMA, "VITIMA"},
  "vitima": {
    "termos": { "VITIMA" },
    "qualidade": {
      "completude": 7, "atualidade": 1,
      "consistência": 0, "relevância": 0 } },
  "criminoso": {
    "termos": { "DOIS INDIVIDUOS", "GARUPA" },
    "qualidade": {
      "completude": 11, "atualidade": 1,
      "consistência": 0, "relevância": 0 } },
  "local": {
    "termos": { },
    "qualidade": {
      "completude": 0, "atualidade": 1,
      "consistência": 0, "relevância": 0 } },
  "objeto": {
    "termos": { "VEICULO", "MOTOCICLETA PRETA", "UMA ARMA" },
    "qualidade": {
      "completude": 55, "atualidade": 1,
      "consistência": 0, "relevância": 0 } }
}
```

Figura 6.1: Denúncia 1 pós Avaliação de Objetos e persistida na Camada de Representação da Informação como Situação e com Índices de Qualidade

Analisando o JSON expresso na figura acima, fica claro a realização das Buscas Contextual e Hierárquica, que realiza a associação de termos complementares, como “MOTOCICLETA PRETA” ou “DOIS INDIVIDUOS”, dois termos compostos que oferecem informações relevantes para construção de SAW sobre esta situação, mas que se analisados de maneira isolada não teriam a mesma expressividade.

Os termos relevantes também já estão relacionados a classes que os identificam, Vítima, Criminoso, Local e Objeto, essa separação é essencial para o processo de Busca Sinérgica realizada antes fusão, pois segmenta as buscas já

identificando mais classes devem ser consideradas. Outra informação presente, são as avaliações de qualidade, tanto nos elementos quanto um índice global de certeza.

Entidades que não apresentam nenhum termo que a classifica, como a classe Local na Figura 6.1, tem o índice Completude com valor "0", pois o cálculo de completude é realizado considerando os elementos identificados. Os demais índices apenas serão consideração após a Associação Multicritério.

```
{
  "termos": {
    "DOIS", "CARAS", "MOTO", "COR", "ESCURA", "ROUBARAM", "UM", "CARRO", "RUA",
    "DA", "MOOCA", "COM", "A", "TAQUARI", "UM", "REVOLVER", "CARRO", "MULHER",
    "CASACO", "VERMELHA"},
  "vitima": {
    "termos": { "MULHER" },
    "qualidade": {
      "completude": 7, "atualidade": 1,
      "consistência": 0, "relevância": 0 } },
  "criminoso": {
    "termos": { "DOIS CARAS" },
    "qualidade": {
      "completude": 6, "atualidade": 1,
      "consistência": 0, "relevância": 0 } },
  "local": {
    "termos": { "RUA DA MOOCA", "COM A", "TAQUARI" },
    "qualidade": {
      "completude": 34, "atualidade": 1,
      "consistência": 0, "relevância": 0 } },
  "objeto": {
    "termos": { "MOTO COR ESCURA", "CARRO", "UM REVOLVER",
      "CASACO VERMELHO" },
    "qualidade": {
      "completude": 55, "atualidade": 1,
      "consistência": 0, "relevância": 0 } }
}
```

Figura 6.2: Denúncia 2 pós Avaliação de Objetos e persistida na Camada de Representação da Informação como Situação e com Índices de Qualidade

```

{
  "termos": {
    "DOIS", "MELIANTES", "ROUBARAM", "UM", "CARRO", "COR", "CINZA", "RUA", "DA",
    "MOOCA", "BANCO", "SANTANDER", "MELIANTES", "MOTO", "DE", "COR", "PRETA",
    "BLUSA", "AZUL", "CALCA", "JEANS", "HOSPITAL", "VILLA", "LOBOS"},
  "vitima": {
    "termos": { },
    "qualidade": {
      "completude": 0, "atualidade": 1,
      "consistência": 0, "relevância": 0 } },
  "criminoso": {
    "termos": { "DOIS MELIANTES", "MELIANTES" },
    "qualidade": {
      "completude": 11, "atualidade": 1,
      "consistência": 0, "relevância": 0 } },
  "local": {
    "termos": { "RUA DA MOOCA", "BANCO SANTANDER",
    "HOSPITAL VILLA LOBOS" },
    "qualidade": {
      "completude": 34, "atualidade": 1,
      "consistência": 0, "relevância": 0 } },
  "objeto": {
    "termos": { "UM CARRO COR CINZA", "MOTO DE COR PRETA", "BLUSA AZUL",
    "CALCA JEANS" },
    "qualidade": {
      "completude": 67, "atualidade": 1,
      "consistência": 0, "relevância": 0 } }
}

```

Figura 6.3: Denúncia 3 pós Avaliação de Objetos e persistida na Camada de Representação da Informação como Situação e com Índices de Qualidade

No processo sintático de fusão, o conteúdo dentro dos JSONs é analisado através de algumas regras, realizando comparações booleanas, se uma palavra é igual a outra, considerando algumas dessas variações, como tensão, gênero e radical, também são construídas algumas regras simples utilizando as classes encontradas. Este processo interno de fusão de informações tem início com a busca de informações sinérgicas, em outras fontes de dados e em situações previamente processadas, que têm algum grau de covariância em sua sintaxe.

Após esta busca sinérgica para outras situações, aqueles que tinham algum grau de sinergia são agrupados em conjunto de dados, usando o exemplo das três

denúncias mencionadas, se não houvesse nenhuma outra informação sinérgica com outra situação seria criado um conjunto apenas com as três informações. Neste caso existem alguns pontos de covariância entre as informações, mesmo que não muito expressivos para a situação como um todo, são suficientes para realizar uma fusão sintática. Termos estes como: “CARRO”, “MOTO”, “DOIS”. Porém mesmo representando os elementos presentes nas denúncias, não indicam um contexto ou informações mais ricas e precisas para fundir estas em uma única situação.

Esses conjuntos ou coleções de informações geradas passam então para uma fase denominada Associação Multicritérios, que considera para a fusão, além dos conjuntos de dados sinérgicos, os índices de qualidade atribuídos aos dados e informações inferidas. Esta atividade organiza e vincula à situação corrente, as novas informações de acordo com a presença de termos e classes, visando sempre a melhoria da qualidade, para que o produto seja uma situação única, completa e a mais precisa possível.

Ao final do processo, o resultado é enviado novamente para a camada de avaliação da qualidade de dados, para ser avaliado e obter novos indicadores de qualidade. O resultado desse processo já com índices de qualidade pode ser observado na Figura 6.4, em formato de diagrama para facilitar a compreensão das informações e comparação com os resultados semânticos. Este resultado é então persistido novamente na camada de Representação da Informação, como uma nova situação, até que seja requerido para a interface do operador ou por um novo processo de fusão.

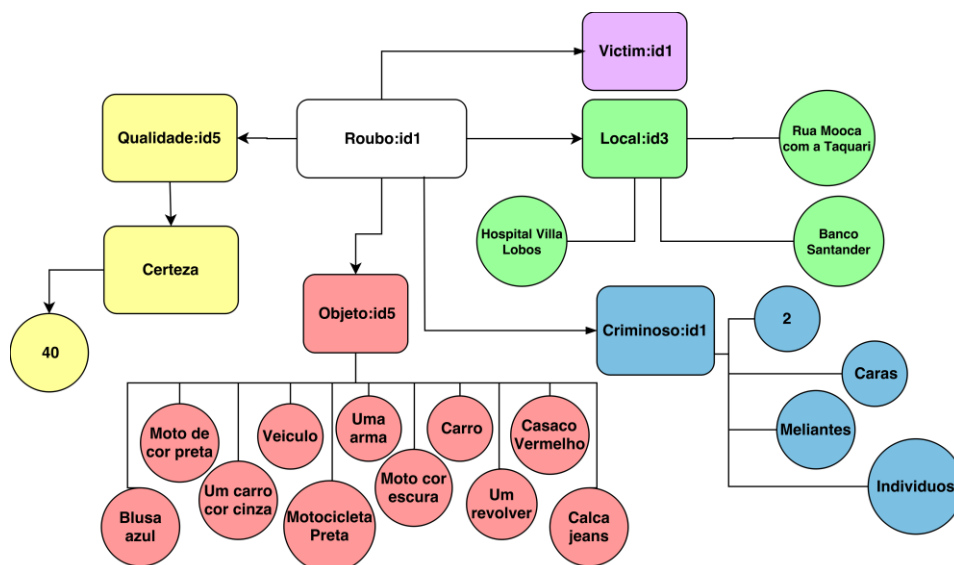


Figura 6.4: Situação Obtida com a Fusão Sintática Durante o Estudo de Caso

Para cada denúncia, são feitas diversas consultas SPARQL na ontologia construída, uma consulta para cada elemento identificado, essa consulta busca identificar qual classe, ou conjunto de classes, da ontologia melhor representa este e as prováveis propriedades que o caracterizam, mesmo que estas não estejam explícitas na denúncia. O resultado desse processo é uma instância de uma situação, que não necessariamente representa um crime.

As Figuras de 6.5 a 6.7, representam o resultado desse processo, descrito acima, respectivamente para as Denúncias 1, 2 e 3. A metodologia para execução das consultas para estruturação desses RDFs, parte das instâncias dos elementos, como se fossem folhas, junto a identificação de seus atributos e propriedades, expressas nas denúncias ou não. Nesse ponto, temos diversas instâncias sem relacionamentos entre elas, porém que pertencem a uma mesma denúncia. O próximo passo é, através de consultas SPARQL, agora locais, identificar propriedades comuns entre as instâncias dos elementos para que seja possível inferir instâncias, propriedades e atributos, com o intuito de definir uma situação para cada denúncia.

Nas figuras, os quadrados representam as instâncias, os círculos os atributos e as setas e traços, as propriedades de relacionamento entre classes e atributos, sendo que cada uma leva seu nome, definido conforme um vocabulário também desenvolvido para o domínio. As instâncias estão divididas por cores, conforme as classes que representam, para facilitar o entendimento da situação, conforme essa separação: rosa, representam objetos relacionados a vítima; vermelho, objetos relacionados ao criminoso; verde, relacionados à local; roxo, são de vítima; azul, de criminoso; branco, representa a instância da situação.

Analisando os elementos encontrados na Denúncia 1, temos os seguintes termos: vítima, veículo, dois indivíduos, moto preta, garupa, arma, ameaça, subtraiu e veículo. Com base nesses termos são feitas consultas na ontologia global, que irá trazer qual classe que se encaixam, ou se são apenas propriedades. Neste caso, um conjunto de classes é retornado com vítima, criminoso e objeto, algumas propriedades como ameaçou e subtraiu, estas representam indícios e sinônimos de um roubo, respectivamente. Com essas buscas SPARQL em uma ontologia rica, também é possível fazer associações entre termos distintos, por exemplo neste caso, não temos o termo “criminoso”, mas uma classe Criminoso é identificado, isso

é possível pois na ontologia um dos termos instanciados que caracteriza um criminoso é “indivíduos”.

E analisando a figura, percebemos uma propriedade “porta” de criminoso para arma de fogo, indicando que havia porte de arma pelos criminosos, porém também não existe um termo “segurava” ou “empunhava” na denúncia, mas sim “com a”, que através do vocabulário é possível estabelecer uma covariância entre estes termos. No final do processo, as instâncias são correlacionadas, conforme as propriedades específicas, encontradas nas denúncias, ou gerais, de dependência entre classes, já previstas no vocabulário. O resultado é demonstrado na Figura 6.5.

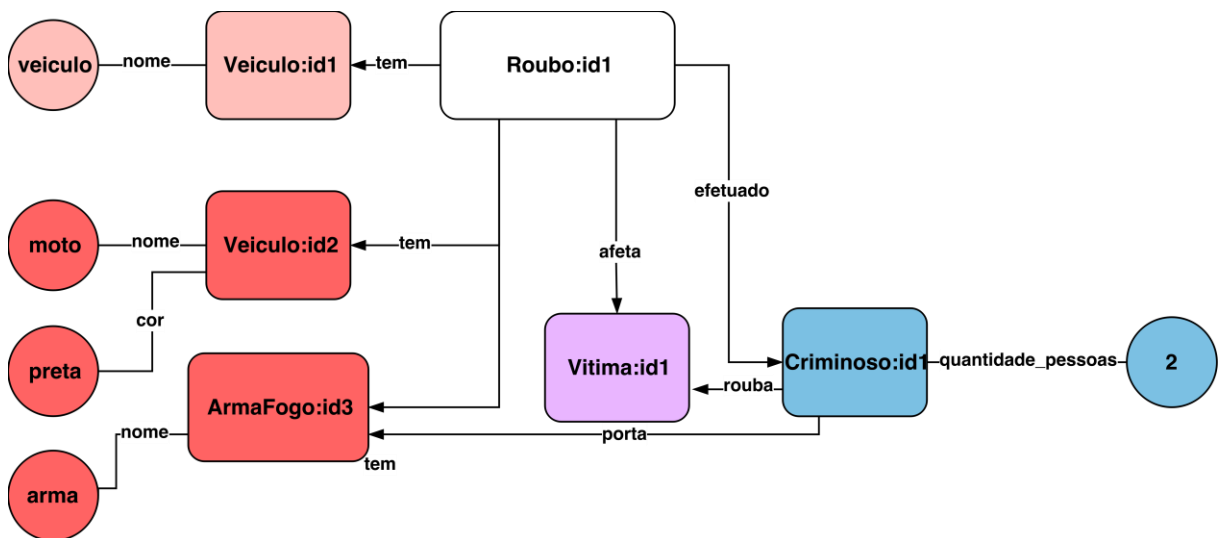


Figura 6.5: RDF da Situação Gerada a Partir da Avaliação de Objetos da Denúncia 1

Neste momento, com a Denúncia 1 totalmente instanciada, conforme a figura acima, onde todas as instâncias dos elementos encontrados na denúncia estão associadas à uma instância maior de situação. Essa informação, já semântica nesse instante é persistida na camada de representação de informação. Esse processo de instanciação dos elementos, inferência interna das propriedades e elaboração da instância de uma única situação com todos os elementos, acontece para todas as denúncias que são submetidas ao processo semântico, nesse caso, três denúncias, o resultado das Denúncias 2 e 3, podem ser vistos nas Figuras 6.6 e 6.7 respectivamente.

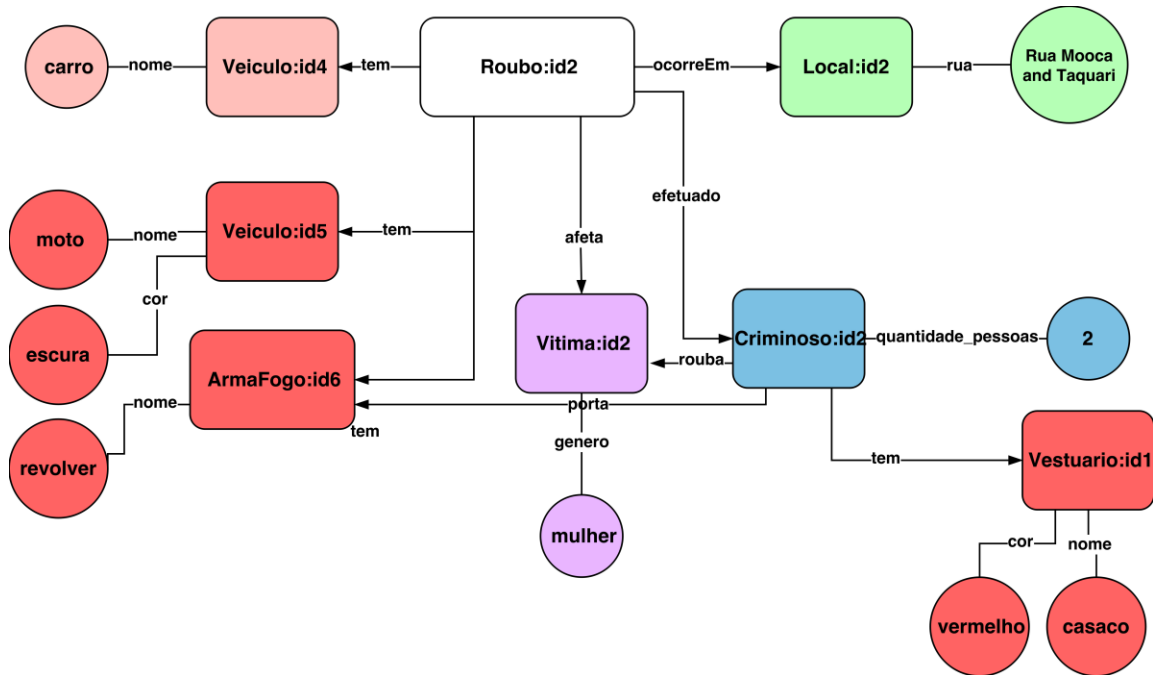


Figura 6.6: RDF da Situação Gerada a Partir da Avaliação de Objetos da Denúncia 2

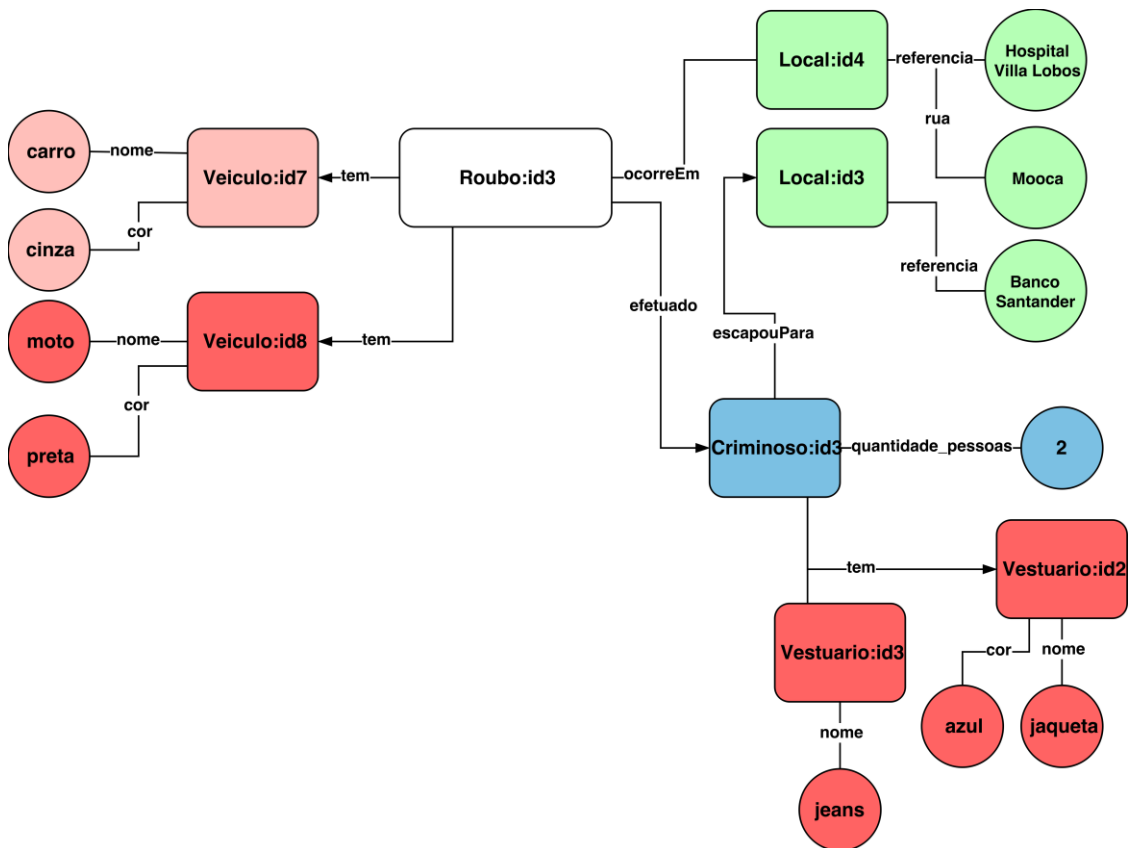


Figura 6.7: RDF da Situação Gerada a Partir da Avaliação de Objetos da Denúncia 3

Como é possível observar as três denúncias parecem ser referentes a uma mesma situação, sendo que, fragmentos do mesmo foram descritos por pessoas diferentes presentes na situação. Essa relação é facilmente compreendida por um

humano, que deduz e transforma todas em um único conhecimento. Porém em cenário mais complexos ou com mais informações humanas se tornam favoráveis a erros e não conseguem absorver todos os possíveis elementos presentes em uma situação ou compreendem um elemento fora de seu contexto real na situação.

Nesse ponto entra a fusão semântica de informações, que a partir das denúncias já processadas, identificação de objetos relevantes e instância de RDFs. É capaz de processar um conjunto de informações gigantes com uma precisão muito alta, e ainda com modelo computacional muito próximo de um modelo mental humano, graças a aplicação das ontologias e vocabulários.

Esse processo de fusão semântico, parte das informações presentes na camada de representação de informações, pós processo de identificação semântica, nesse ponto, o processo é muito parecido com a parte de estruturação das denúncias, todavia as inferências são feitas entre as instâncias de todas as denúncias, analisando todas as propriedades presentes em cada instância interna dos elementos e fazendo associações entre eles. No final deste processo, teremos um novo conjunto de possíveis situações variando na organização e presença dos elementos e propriedades.

Cada uma dessas possíveis situações é salva na camada de representação e tem sua qualidade avaliada, nesse ponto são executadas as avaliações multicritério, principalmente avaliando as melhoras nos índices de qualidade. Aquela que apresentar a melhor qualidade e que satisfaz outros critérios impostos, como a presença de algum elemento específico, será eleita como situação final resultante da fusão e persistida na camada de representação e futuramente apresentada na interface.

A fusão dessas informações permite a obtenção de uma informação com maior valor agregado pela junção dos fragmentos. Considerando essa situação de estudo, com as três denúncias, descritas acima, a informação fundida permite a identificação da vestimenta de ambos os criminosos, características do objeto roubado, arma utilizada pelos criminosos e veículo de fuga, como também o local onde ocorreu o crime, informações que de maneira isolada não apresentavam tanto valor, o resultado desse processo é demonstrado na Figura 6.8.

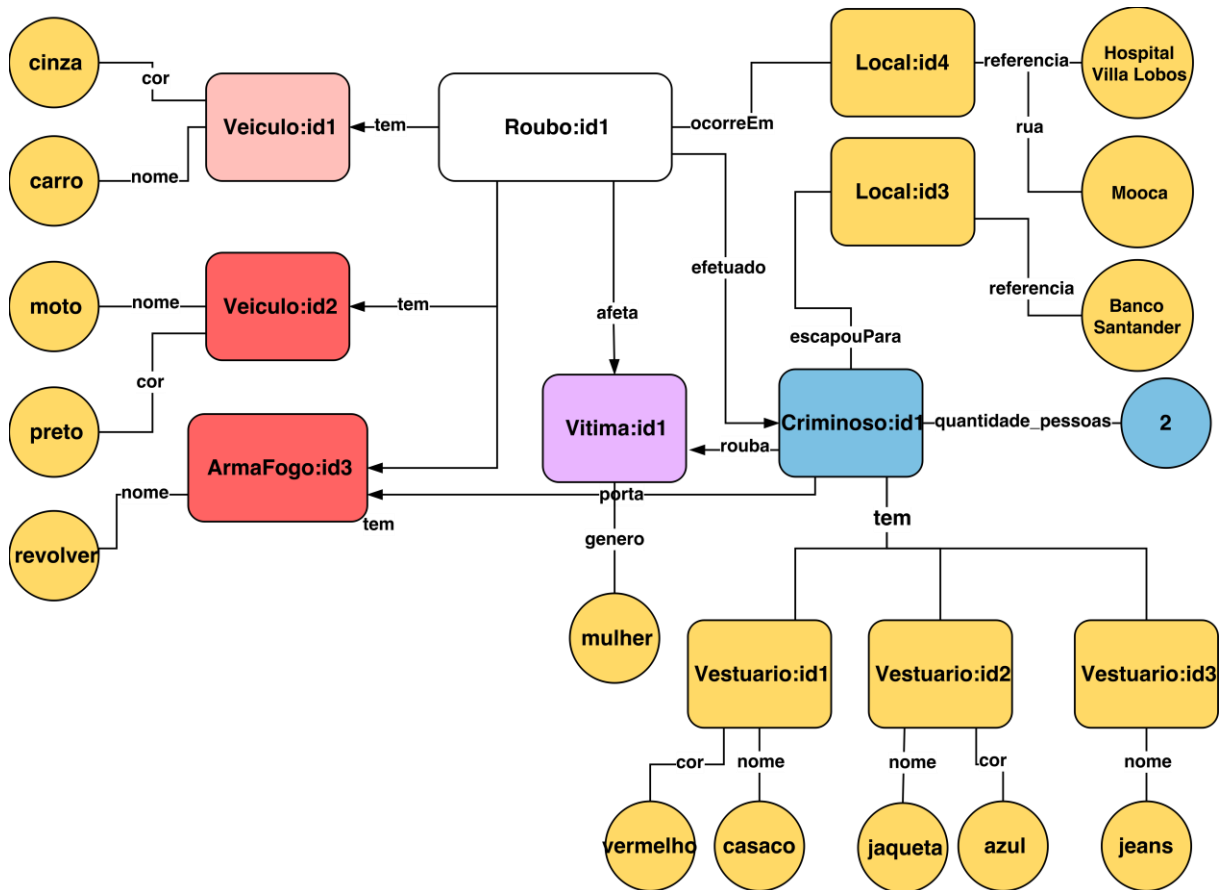


Figura 6.8: Resultado do Processo Semântico de Busca e Integração de Informações Sinérgicas com Informações do Estudo de Caso

Como na ilustração do processo de análise semântica de cada denúncia, a Figura 6.8 que representa o resultado da fusão semântica também está dividida por cores. A única diferença entre as cores já descritas anteriormente, é a cor amarela que representa as informações que foram fundidas entre as três denúncias. Além da presença de mais instâncias e atributos na situação fundida, também é possível notar novas propriedades entre os elementos, que analisando as denúncias separadamente não eram explícitas para a situação.

Um outro ponto a ser observado é que a qualidade da informação pode ser aplicada na ontologia para permitir verificar quais informações possuem uma melhor qualidade, e assim decidir o que deve ser utilizado na fusão ou não. Para isso é possível utilizar a ontologia *Data Quality Vocabulary* (DQV) que é construída em cima de outra ontologia de qualidade DAQ.

DQV permite a criação de instâncias de categorias, dimensões e métricas para informar a qual uma determinada medida de qualidade se refere. Para uma determinada instância da ontologia é criada uma instância de *QualityGraph*, a

mesma faz a conexão entre todas as medidas de qualidade aplicadas para aquela informação instanciada. Um exemplo é no caso de um roubo, nesse caso a instância de roubo estará relacionada a uma instância de *QualityGraph*, que por si só, estará relacionada a todas as instâncias das medidas de qualidade, exemplificado na Figura 6.9.

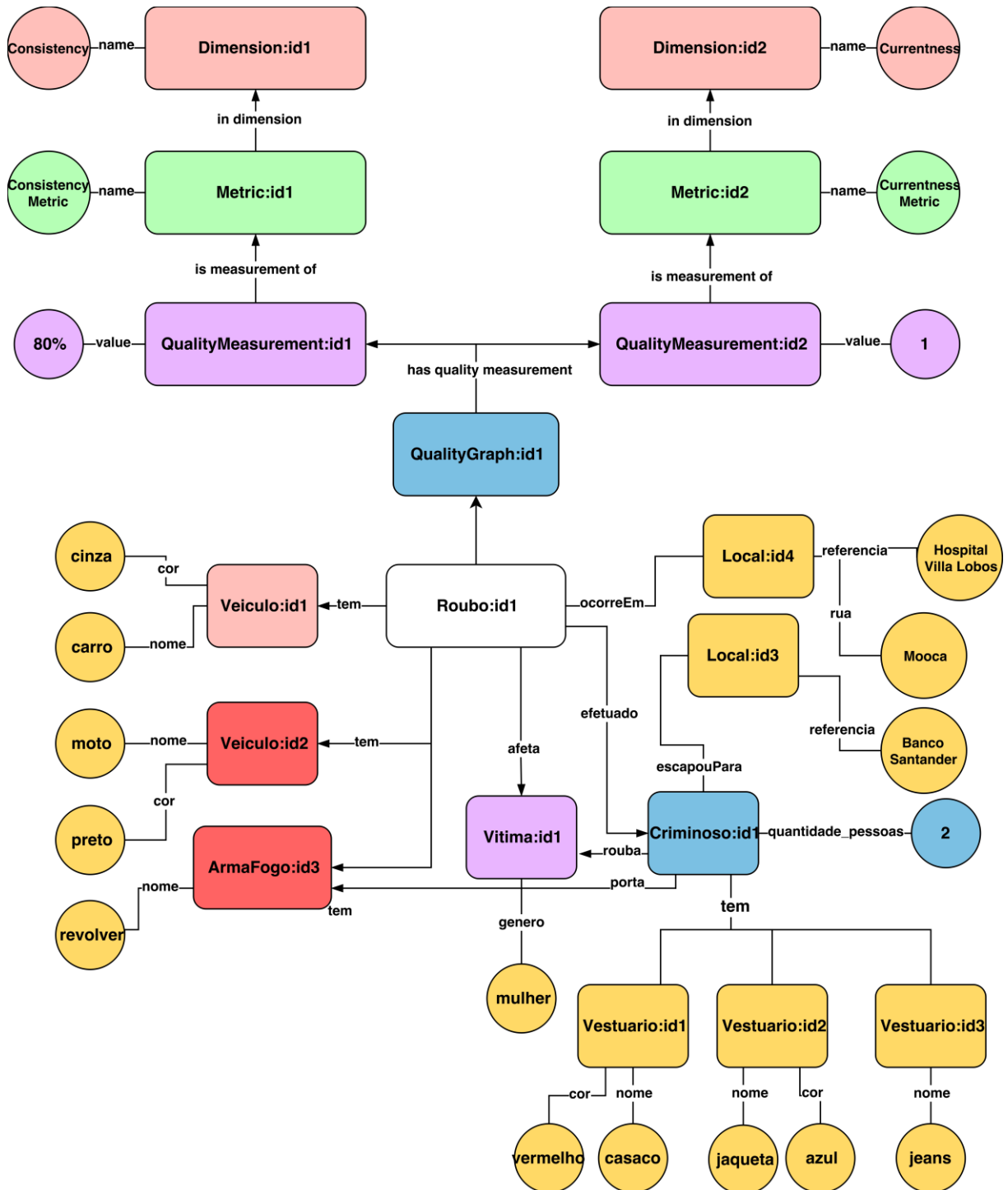


Figura 6.9: Resultado do Processo Semântico de Busca e Integração de Informações Sinérgicas Aplicado o DQV para Representar a Qualidade

Capítulo 7

CONCLUSÕES

O presente trabalho tem como objetivo o desenvolvimento de um processo de fusão de informações para a melhoria da qualidade da informação em situação de risco, por meio de abordagens sintática e semântica e auxílio de módulos de avaliação de qualidade e ontologias para representação de informações. Os benefícios atingidos com este trabalho é principalmente a melhoria da qualidade final das informações, com uma dimensão menor, porém com mais representatividade, que por sua vez afeta de maneira positiva o processo de construção de SAW e tomada de decisão de operadores.

O trabalho demonstrou o desenvolvimento de técnicas semânticas durante o processo de fusão de informações, seguindo as diretrizes do modelo Quantify, utilizando de módulos de avaliação de objetos e situações. Essa abordagem junto à implementação sintática, realizando diversos processos de tratativa e avaliação de qualidade nas informações processadas e geradas, demonstra alta possibilidade de sucesso ao ser implantada em rotinas de avaliação de situações reais, auxiliando Sistemas de Gerenciamento de Risco e Apoio a Tomada de Decisão a adquirir informações melhores, possibilitando assim que especialistas possam desenvolver SAW com informações mais concisas, precisas e com índices de qualidade melhores.

As avaliações de qualidade em suporte ao processo possibilitaram a análise de diversos resultados parciais e finais, demonstram o quão útil se faz a adoção dos processos de fusão, tanto sintático quanto semântico, apresentando informações relevantes para a construção de SAW, de acordo com os requisitos definidos pelos especialistas de domínio. Todavia ainda é um domínio de estudo muito complexo, sendo que este trabalho apresenta resultados apenas comprobatórios de que

adoção de processo de fusão, análise de qualidade e semântica se fazem uteis para a colaboração com humanos.

A aquisição de SAW está diretamente relacionada com a qualidade da informação e a fusão de informações se propõe a gerar informações de menor dimensão e de qualidade superior, fornecendo melhores subsídios para SAW e conseqüentemente para a tomada de decisão. Todavia essa análise dos resultados será realizada em trabalhos futuros baseada em uma avaliação dos resultados gerados pelo processo de fusão proposto, ao processar dados reais provenientes de registros presentes em bases oficiais de situações de crime, especificamente roubo e furto, junto a aplicação de alguma metodologia para avaliação de SAW, comprovando de maneira definitiva os ganhos ao aplicar o processo aqui descrito.

É claro que restam muitas incertezas e falhas no processo quando considerado a imprevisibilidade do domínio, uma estrutura ontológica que suporte e possa representar qualquer informação que seja recebida, assim como um conjunto de regras de associação que possa atender essa ontologia são processos complexos. Onde a adesão de outras técnicas de inteligência e manipulação de dados podem ser úteis, para técnicas avançado de Aprendizado de Máquina e Inteligência Artificial.

Dessa maneira é possível concluir que a implantação da semântica em processos de Fusão de Dados e Informações, demonstra grandes avanços promissores as linhas de pesquisa e soluções voltadas o domínio de gerenciamento de risco e informações. Também ficou claro o quão relevante se torna para processos de SAW em domínio críticos, conseguindo gerar um nível muito mais alto de informação para um operador, com resultados promissores mesmo aplicado em ambientes controlados.

7.1 Publicações Obtidas

- *Fusão de dados de fontes heterogêneas para sistema de tomada de decisão militar*. Valdir Junior, João Ricardo, Leonardo Botega. 14º Congresso Nacional de Iniciação Científica. São Paulo, São Paulo, Brasil, 2014;

-
- *Fusão de dados de fontes heterogêneas para sistema de tomada de decisão militar.* Valdir Junior, Leonardo Botega. 4º Congresso de Pesquisa Científica: Inovação, Sustentabilidade, Ética e Cidadania. Marília, São Paulo, Brasil, 2014;
 - *Fusão semântica multicritério de informações de fontes heterogêneas com índices de qualidade em sistemas de suporte a emergência.* Valdir Junior, Leonardo Botega. 5º Congresso de Pesquisa Científica: Inovação, Sustentabilidade, Ética e Cidadania. Marília, São Paulo, Brasil, 2015;
 - *Multi-criteria Fusion of Heterogeneous Information for Improving Situation Awareness on Emergency Management Systems.* Valdir Junior, Matheus Sanches, Leonardo Botega, Jessica Souza, Caio Saraiva Coneglian, Elvis Fusco, Márcio Roberto de Campos. 17th International Conference on Human-Computer Interaction. Los Angeles, Estados Unidos, 2015;
 - *Fusão Semântica utilizando Qualidade da Informação e a Avaliação de Objetos e Situações para Melhoria da Consciência Situacional em Situações de Emergência.* Valdir Junior, Leonardo Botega. 6º Congresso de Pesquisa Científica: Inovação, Sustentabilidade, Ética e Cidadania. Marília, São Paulo, Brasil, 2016;
 - *Objects Assessment Approach Using Natural Language Processing and Data Quality to Support Emergency Situation Assessment.* Matheus Sanches, Valdir Junior, Jessica Souza, Caio Saraiva Coneglian, Fábio Jorge, Natália Oliveira, Leonardo Botega. 18th International Conference on Human-Computer Interaction. Toronto, Canadá, 2016.
 - *Ontological Semantic Agent in the Context of Big Data: A Tool Applied to Information Retrieval in Scientific Research.* Caio Saraiva Coneglian, Elvis Fusco, José Eduardo Santarém, Valdir Junior, Leonardo Castro Botega. Advances in Intelligent Systems and Computing. 2016;
 - *Development of a User Interface for the Enrichment of Situational Awareness in Emergency Management Systems.* Natália Oliveira, Fábio Jorge, Jessica Souza, Valdir Junior, Leonardo Botega. 7th

International Conference on Applied Human Factors and Ergonomics. Orlando, Florida, Estados Unidos, 2016;

- *Using Semantics to Improve Information Fusion and Increase Situational Awareness.* Valdir Junior, Matheus Sanches, Leonardo Botega, Caio Saraiva Coneglian, Natália Oliveira, Regina Araújo. 7th International Conference on Applied Human Factors and Ergonomics. Orlando, Florida, Estados Unidos, 2016;
- *Towards Semantic Fusion Using Information Quality Awareness to Support Emergency Situation Assessment.* Valdir Junior, Matheus Sanches, Leonardo Botega, Caio Saraiva Coneglian. 4th World Conference on Information Systems and Technologies. Recife, Pernambuco, Brasil, 2016;
- *Towards semantic fusion using information quality and the assessment of objects and situations to improve emergency situation awareness.* Valdir Junior, Matheus Sanches, Jordan Saran, Caio Saraiva Coneglian, Leonardo Botega. 11th International Conference on Digital Information Management. Porto, Portugal, 2016;
- *Data and information fusion in the context of emergency management: The DF100Fogo project.* Jordan Saran, Vitória Mendes, Valdir Junior, Cassio Santos, Gabriel Nascimento, Maria de Fatima Tavares, Allan Oliveira, Leonardo Botega. 12th Iberian Conference on Information Systems and Technologies. Lisboa, Portugal, 2017;
- *Beyond syntactic data fusion in the context of criminal data analysis.* Valdir Junior, Jordan Saran, Lucas Zanco, João Martins, Vagner Pagotti, Allan Souza, Leandro Villas, Leonardo Botega. 12th Iberian Conference on Information Systems and Technologies. Lisboa, Portugal, 2017;
- *Quality-aware human-driven information fusion model.* Leonardo Botega, Valdir Junior, Allan Oliveira, Jordan Saran, Leandro Villas, Regina Araújo. 20th International Conference on Information Fusion. Xian, China, 2017.

REFERÊNCIAS

BAADER, F., HORROCKS, I. SATTLER, U., Description logics as ontology languages for the semantic web. In: **Mechanizing Mathematical Reasoning**, p. 228-248, 2005.

BATINI, C., CAPPIELLO, C., FRANCALANCI, C., MAURINO, A. Methodologies for data quality assessment and improvement. In: **ACM Computing Surveys**, v. 41, no. 3, p. 1-52, julho, 2009.

BLASCH, E. P., LAMBERT, D. A., VALIN, P., KOKAR, M. M., LLINAS, J., DAS, S., CHONG, C. & SHAHBAZIAN, E. High level information fusion (HLIF): Survey of models, issues, and grand challenges. In: **IEEE Aerospace and Electronic Systems Magazine**, v. 27, n.9, p. 4-20, 2012.

BLASCH, E., STEINBERG, A., DAS, S., LLINAS, J., CHONG, C., KESSLER, O., WALTERS, E., WHITE, F., Revisiting the JDL model for information Exploitation. In: **16th International Conference on Information Fusion**, p. 129-136, 2013.

BLASCH, E., VALIN, P., JOUSSELME, A. L., LAMBERT, D., & BOSSÉ, É., Top ten trends in high-level information fusion. In: **15th International Conference on Information Fusion**, p. 2323-2330, 2012.

BOTEGA, L. C., BERTI, C., ARAUJO, R., ALMEIRA NERIS, V. P., A Model to Promote Interaction between Humans and Data Fusion Intelligence to Enhance Situational Awareness, Human-Computer Interaction. Theories, Methods, and Tools. In: **International Conference on Human-Computer Interaction**, pp.399-410, 2014.

BOTEGA, L. C., Modelo de Fusão Dirigido por Humanos e Ciente de Qualidade de Informação. Universidade Federal de São Carlos, Centro de Ciências Exatas e de Tecnologia, Programa de Pós-Graduação em Ciência da Computação, 2015.

BOTEGA, L. C., SOUZA, J. O., JORGE, F. R., CONEGLIAN, C. S., CAMPOS, M. R., ALMEIDA NERIS, V. P., & ARAÚJO, R. B. Methodology for data and information quality assessment in the context of emergency situational awareness. In: **Universal Access in the Information Society**, p. 1-14, 2016.

CARVALHO, R. N., MATSUMOTO, S., LASKEY, K. B., COSTA, P. C., LADEIRA, M., & SANTOS, L. L. Probabilistic ontology and knowledge fusion for procurement fraud detection in brazil. In: **Uncertainty Reasoning for the Semantic Web II**, p. 19-40, 2013.

CASTANEDO, F. A review of data fusion techniques. In: **The Scientific World Journal**, v. 2013, 2013.

COSTA, P. C., LASKEY, K. B., CHANG, K. C., SUN, W., PARK, C. Y., MATSUMOTO, S., High-level information fusion with bayesian semantics. In: **Proceedings of the 9th Bayesian Modelling Applications Workshop**, 2012.

ENDSLEY, M. R., Design and Evaluation for Situation Awareness Enhancement. In: **Proceedings of the human factors and ergonomics society annual meeting**, v. 32, n. 2, p. 97-101, 1988.

ENDSLEY, M. R., Designing for situation awareness in complex systems, Proceedings of the Second International Workshop on symbiosis of humans, artifacts and environment. In: **Proceedings of the Second international workshop on symbiosis of humans, artifacts and environment**, novembro 2001.

ENDSLEY, M. R., JONES, D. G., Designing for Situation Awareness: an approach to user-centered design, second edition, Taylor & Francis, 2012.

FOO, P., NG, G. High-level Information Fusion: an overview. In: **Journal of Advanced Information Fusion**, v. 8, n. 1, 2013.

GABA, D. M., HOWARD, S. K., SMALL, S. D., Situation awareness in anesthesiology. In: **Human factor**, v. 37, n. 1, p. 20-31, 1995.

GRUBER, T. R., A translation approach to portable ontology specifications. In: **Knowledge acquisition**, v. 5, n. 2, p. 199-220, 1993.

HALL, D. L., JORDAN, J., Human-centered information fusion, Artech House, 2010.
HALL, D. L., MCMULLEN, S. A. H., Mathematical Techniques in Multisensor Data Fusion, Artech House, 2004.

HALL, D. L., MCNEESE, M., YEN, J., EL-NASAL, M. S., A Three Pronged Approach for Improved Data Understanding: 3-d visualization, use of gaming techniques, and inteligente advisory agentes. Pennsylvania State Univ University., Park, Pa. College Of Information Sciences and Technology., 2006.

JONES, R. E.; CONNORS, E. S.; ENDSLEY, M. R. Incorporating the human analyst into the data fusion process by modeling situation awareness using fuzzy cognitive maps. In: **12th International Conference on Information Fusion**, p. 1265-1271, 2009.

JUNIOR, V. A. P., SANCHES, M. F., BOTEGA, L. C., & CONEGLIAN, C. S. Towards semantic fusion using information quality awareness to support emergency situation assessment. In: **New Advances in Information Systems and Technologies**. Springer, Cham, p. 145-155, 2016.

KESSLER, J., Functional Description of the Data fusion Process – Presentation to the US Naval Board. Office of Naval Technology Data Fusion Development Strategy, 1991.

KIM, H., SON, J., JANG, K., Semantic Data Fusion: from Open Data to Linked Data. In: **Proceedings of the European Semantic Web Conference**, 2013.

KOKAR, M. M., ENDSLEY, M. R., Situation Awareness and Cognitive Modeling. **IEEE Intelligent Systems**, v. 27, n. 3, p. 91-96, maio 2012.

LASKEY, K., NG, G. W., NAGI, R., SUNY-BUFFALO, C. M. I. F., STAMPOULI, D., SCHUBERT, J., & VALIN, P. Issues of Uncertainty Analysis in High Level Information Fusion. In: **Int'l Conf. On Information Fusion**, 2012.

LLINAS, J., BOWMAN, C., ROGOVA, G., STEINBERG, A., WALTZ, E., WHITE, F. Revisiting the JDL data fusion model II. In: **Space and Naval Warfare Systems Command Sand Diego CA**, 2004.

MCGUINNESS, D. L., Ontologies for information fusion. In: **Proceedings of the Sixth International Conference on Information Fusion**, p. 650-656, julho 2003.

NILSSON, M., LAERE, J. V., SUSI, T., ZIEMKE, T., Information fusion in practice: A distributed cognition perspective on the active role of users. In: **Information Fusion**, v. 13, p. 60-78, janeiro 2012.

NOY, N. F., MCGUINNESS, D. L., Ontology development 101: A guide to creating your first ontology, 2001.

PEREIRA, V. A., SARAN, J. F., LADEIRA, L. Z., MARTINS, J. H., PAGOTTI, V., SOUZA, A. M., VILLAS, L. A., BOTEAGA, L. C. Beyond syntactic data fusion in the context of criminal data analysis. In: **12th Iberian Conference on Information Systems and Technologies (CISTI)**, p.1, 2017.

ROGOVA, G., BOSSE, E. Information quality in information fusion. In: **13th Information Fusion**, 2010.

ROGOVA, G., NIMIÉR, V., Reliability in information fusion: literature survey. In: **7th International Conference on Information Fusion**, v. 2, p. 1158-1165, 2004.

SALERNO, J., Information fusion: a high-level architecture overview. In: **Proceedings of the Fifth International Conference on Information Fusion**, v. 1, p. 680-686, 2002.

SANCHES, M. F., JUNIOR, V. A. P., SOUZA, J. O., CONEGLIAN, C. S., JORGE, F. R., OLIVEIRA, N. P., BOTEAGA, L. C. Objects Assessment Approach Using Natural Language Processing and Data Quality to Support Emergency Situation Assessment. In: **International Conference on Human-Computer Interaction**. Springer International Publishing, p. 238-244, 2016.

SHELTON, C. L., KINSTON, R., MOLYNEUX, A. J., AMBROSE, L. J. Real-time situation awareness assessment in critical illness management: adapting the situation present assessment method to clinical simulation. **BMJ quality & safety**, 2012.

SOUZA, J. BOTEAGA, L., SANTARÉM SEGUNDO, J.E., BERTI, C., Conceptual Framework to Enrich Situation Awareness of Emergency Dispatchers, Human Interface and the Management of Information. Information and Knowledge. In:

International Conference on Human Interface and the Management of Information, p. 33-44, 2015.

STANTON, N. A., CHAMBERS, P. R., PIGGOTT, J., Situational awareness and safety. In: **Safety Science**, v. 39, n. 3, p. 189-204, 2001.

STEINBERG, A, N., BOWMAN, C. L., WHITE, F. E., Revisions to the JDL data fusion model. In: **Proceedings of SPIE**, v.3719, n.1, p. 430-441, 1999.

STEINBERG, A., N., BOWMAN, C. L., Chapter 2: revisions to the jdl data fusion model. In: David L. Hall and James Llinas (Ed.). **Multisensor Data Fusion**. CRC Press, 2001.

VINCEN, D., STAMPOULI, D., POWEL, G. Foundations for system implementation for a centralised intelligence fusion framework for emergency services. In: **12th International Conference on Information Fusion**, p. 1401-1408, 2009.

WHITE JR, E. F., JDL, Data Fusion Lexicon. In: TECHNICAL PANEL FOR C3, 1991, San Diego, Calif, USA. **Anais...**, 1991.

SPORNY, M.; KELLOGG, G.; LANTHALER, M. JSON-LD 1.0: A JSON-based Serialization for Linked Data. Rio de Janeiro: W3C, 2014.

ISOTANI, S., & BITTENCOURT, I. I. Dados Abertos Conectados: Em busca da Web do Conhecimento. Novatec Editora, 2015.