

CENTRO UNIVERSITÁRIO EURÍPIDES DE MARÍLIA
FUNDAÇÃO DE ENSINO “EURÍPIDES SOARES DA ROCHA”
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

**PROCESSAMENTO E ENTENDIMENTO DE
LINGUAGEM NATURAL NO GERENCIAMENTO
DE EMERGÊNCIAS PARA OBTENÇÃO DE
CONSCIÊNCIA SITUACIONAL**

MATHEUS FERRARONI SANCHES

ORIENTADOR: PROF. DR. LEONARDO CASTRO BOTEGA

Marília - SP
Dezembro/2017

CENTRO UNIVERSITÁRIO EURÍPIDES DE MARÍLIA
FUNDAÇÃO DE ENSINO “EURÍPIDES SOARES DA ROCHA”
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

**PROCESSAMENTO E ENTENDIMENTO DE
LINGUAGEM NATURAL NO GERENCIAMENTO
DE EMERGÊNCIAS PARA OBTENÇÃO DE
CONSCIÊNCIA SITUACIONAL**

MATHEUS FERRARONI SANCHES

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
Dezembro/2017



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

BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO


Matheus Ferraroni Sanches


Processamento e Entendimento de Linguagem Natural no Gerenciamento de Emergências
Para Obtenção de Consciência Situacional

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 

1º.Examinador: Fabio Piola Navarro 

2º.Examinador: Guilherme Rodrigues Bilar 

Marília, 27 de novembro de 2017.

RESUMO

A consciência situacional (SAW) envolvendo situações de emergências criminais é um elemento decisivo para a tomada de decisão. Para tal, o tomador de decisão, utilizando um sistema de apoio à decisão, precisa ter em mãos as informações mais relevantes e confiáveis possíveis. A tomada de decisão não fundamentada em informações relevantes pode resultar em uma limitada SAW e conseqüentemente em danos graves a propriedades e/ou a vida. Para auxiliar o processo de aquisição de SAW, este trabalho apresenta um processo para se obter, analisar e entender a linguagem natural de um agente humano em campo com o objetivo de representar semanticamente informações relevantes e melhor classificadas sobre as situações que estes presenciam. Durante este processo de análise, ferramentas como um parser, redes neurais e ontologias são utilizadas em rotinas complementares para a classificação do linguagem. Como resultado, após as análises dos dados simulados de agentes em campo, foi possível gerar uma representação semântica dos elementos relevantes contidos nos dados obtidos. Com esta representação semântica é possível fornecer subsídio para a geração de SAW com qualidade.

Palavras-chave: Processamento de linguagem natural, redes neurais artificiais, ontologia, consciência situacional, gerência de emergência, tomada de decisão

ABSTRACT

A consciência situacional (SAW) envolvendo situações de emergências criminais é um elemento decisivo para a tomada de decisão. Para tal, o tomador de decisão, utilizando um sistema de apoio à decisão, precisa ter em mãos as informações mais relevantes e confiáveis possíveis. A tomada de decisão não fundamentada em informações relevantes pode resultar em uma limitada SAW e conseqüentemente em danos graves a propriedades e/ou a vida. Para auxiliar o processo de aquisição de SAW, este trabalho apresenta um processo para se obter, analisar e entender a linguagem natural de um agente humano em campo com o objetivo de representar semanticamente informações relevantes e melhor classificadas sobre as situações que estes presenciam. Durante este processo de análise, ferramentas como um parser, redes neurais e ontologias são utilizadas em rotinas complementares para a classificação do linguagem. Após a conclusão das análises apresentadas neste trabalho, como resultado, é gerado a partir dos dados obtidos de agentes em campo, uma representação semântica dos elementos relevantes contidos dentro dos dados obtidos. Com esta representação semântica é possível fornecer subsídio para a geração de SAW com qualidade.

Keywords: Natural language processing, artificial neural networks, ontology, situation awareness, emergency management, decision making

LISTA DE FIGURAS

Figura 2.1 – Modelo de Consciência Situacional	20
Figura 2.2 – Exemplo de ontologia	25
Figura 2.3 – Exemplo de rede neural	32
Figura 3.1 – Fluxo Geral das Análises	39
Figura 3.2 – Análise Sintática	40
Figura 3.3 – Análise sintática e gramatical	42
Figura 3.4 – Análise utilizando rede neural artificial	45
Figura 3.5 – Análise utilizando ontologia como apoio	47
Figura 4.1 – Seleção de Técnica para entendimento de linguagem natural	57

LISTA DE TABELAS

Tabela 4.1 – Comparação Entre Técnicas para Realização de Entendimento de Fala Natural	56
--	----

LISTA DE ABREVIATURAS E SIGLAS

SAW - *Situation Awareness*

HUMINT - *Human Intelligence*

NLP - *Natural Language Processing*

NLU - *Natural Language Understanding*

TC - *Text Classification*

GIHC - *Grupo Interação Humano-Computador*

RDF - *Resource Description Framework*

SUMÁRIO

1	INTRODUÇÃO	9
1.1	Motivação	11
1.2	Problema de Pesquisa	11
1.3	Justificativa	13
1.4	Objetivo Geral	13
1.5	Objetivos Específicos	13
1.6	Metodologia	14
1.7	Organização	15
2	CONSCIÊNCIA SITUACIONAL E FERRAMENTAS	16
2.1	Consciência Situacional	16
2.2	Conceitos	16
2.3	Erros de SAW	18
2.3.1	Atualização de SAW	19
2.4	Obtenção de SAW	19
2.4.1	Manutenção de SAW	22
2.5	Objetivos de SAW	23
2.6	Representações de Informações para SAW	23
2.7	Processamento de linguagem natural	27
2.8	Ferramentas Utilizadas para NLP	28
2.8.1	Google Speech To Text	28
2.8.2	Cogroo	28
2.9	Natural Language Understanding (NLU)	29
2.10	Classificação de texto	29
2.11	Obtenção de dados para SAW através de classificação de texto	30
2.12	Inteligência artificial	30
2.13	Redes Neurais Artificiais para Identificação de Elementos Relevantes	31
2.13.1	Ferramentas em Suporte a Rede Neural Artificial	34
2.13.1.1	Encog	34
2.13.2	Redes neurais e processamento de linguagem natural	34
3	PROCESSAMENTO E ENTENDIMENTO DE LINGUAGEM NATURAL NO GERENCIAMENTO DE EMERGÊNCIAS PARA OBTENÇÃO DE CONSCIÊNCIA SITUACIONAL	36
3.1	Trabalhos Correlatos	36

3.2	Desenvolvimento	38
3.3	Identificação de Elementos com Análise Sintática	40
3.4	Gramática para Apoiar a Busca de Elementos Relevantes	41
3.5	Desenvolvimento de Rede Neural para Apoiar a Busca de Elementos Relevantes	43
3.6	Ontologia para Apoiar a Busca de Elementos Relevantes	46
3.7	Continuação Pós Identificação de Elementos	47
	4 ESTUDO DE CASO	49
4.1	Resultados da Análise Sintática	50
4.2	Resultados da Análise Sintática em Conjunto com a Análise Gramatical	51
4.3	Resultados da Rede Neural Com Dados Sintáticos e Gramaticais	53
4.4	Resultados da Rede Neural com Dados Sintáticos, Gramaticais e Semânticos	54
4.5	Considerações Finais	56
	5 CONCLUSÕES	59
5.1	Publicações	60
	REFERÊNCIAS BIBLIOGRÁFICAS	63
	Apêndice A TREINO UTILIZADO NA REDE NEURAL ARTIFICIAL	65
A.1	Arma	65
A.2	Criminoso	67
A.3	Objeto	69
A.4	Tempo	70
A.5	Vitima	72
A.6	Variedades	75

Capítulo 1

INTRODUÇÃO

Consciência Situacional (SAW) é um processo cognitivo gerado pelo entendimento de uma determinada situação. SAW consegue identificar e entender os elementos presentes dentro da situação que está sendo analisada, bem como a interação entre estes elementos. SAW é dividida por níveis onde quanto mais alto for o nível, mais completo e preciso é o modelo gerado por SAW .

Dentro do domínio de tomada de decisão, SAW tem a sua relevância aumentada, já que é de suma importância que se tenha conhecimento da situação para a tomada de decisão correta. A qualidade das ações está diretamente relacionado com a qualidade de SAW, uma vez que pouco conhecimento sobre uma situação gera SAW de baixa qualidade e resulta em problemas de alocação de recurso.

Para a tomada de decisão a qualidade das ações influencia nos resultados obtidos de cada decisão, de forma que decisões incorretas podem causar prejuízo ao patrimônio e a vida. Além disso, deve ser levado em consideração que as situações acontecem no mundo real e em ambientes dinâmicos que precisam ser tratados de maneira exclusiva de modo a manter a atualidade.

A manutenção de SAW é necessária quando situações dinâmicas estão presentes, o que é essencialmente o universo de tomada de decisões com base de dados criminal. Durante a obtenção dos dados necessários para decidir que algo está acontecendo, os elementos envolvidos já estão em estágios diferentes, pois, estão continuamente interagindo com outros elementos. Caso a manutenção de SAW não seja feita em situações dinâmicas, o modelo contido em SAW fica desatualizado, e quando a tomada de decisão for feita, poderá se basear em informações ultrapassadas e incorretas quando comparadas com o momento atual da situação. A realização da tomada de decisão baseada em SAW desatualizado pode resultar em inconsistências e causar prejuízos.

Dentro do domínio de gerenciamento de emergências criminais, duas das prin-

principais fontes de informações sobre determinada situação são as denúncias realizadas através do serviço de resposta a emergências e dados presentes em redes sociais. Essas fontes podem ser utilizadas em tempo real para criar uma nova situação ou aprimorar os dados existentes sobre uma situação, permitindo assim a manutenção do conhecimento sobre esta situação. Cada um destes meios possui maneiras distintas de serem obtidos e maneiras distintas de serem processadas para que se possa, apenas depois de um pré processamento, realizar a busca por elementos relevantes ao domínio nestes dados.

A alta disponibilidade dos dados provenientes de humanos (Human Intelligence -HUMINT) possibilita a obtenção rápida de informações sobre uma determinada situação com uma baixa taxa de atraso. Pessoas distintas reportando a mesma situação com informações diferentes, devido ao que cada pessoa pode ter presenciado, é algo que também auxilia na completude das informações para a geração de SAW.

O problema com a utilização de dados HUMINT acontece quando os dados podem conter inconsistências e incoerências com a realidade, uma vez que durante a denúncia a pessoa que está relatando o acontecido por sofrer de fatores externos e omitir ou alterar informações importantes. Eventos que acontecem com a pessoa que esta disponibilizando as informações podem fazer com que ela mude subitamente o que esteja falando, pare de falar e/ou até mesmo crie informações que não existam de verdade, por medo ou estresse.

Com o objetivo de procurar elementos relevantes para a geração de SAW de qualidade este trabalho propõe técnicas distintas para a realização da análise destes elementos, de forma que o resultado tenha a maior qualidade possível.

Como o processamento de dados HUMINT é algo de grande complexidade, para a realização dessa tarefa foi realizado o uso de processamento de linguagem natural (NLP) com o objetivo de entender o que esta sendo transmitido dentro do texto (NLU). Para que fosse possível realizar NLU, foram utilizadas técnicas que consideram a sintática, gramática e semântica das palavras, que foram classificadas utilizando algoritmos de aprendizado de máquina.

Com o uso e documentação das técnicas utilizadas, pode ser observado onde e quando cada uma das técnicas utilizadas pode apresentar vantagens, assim foi definido qual o melhor conjunto de técnicas e ferramentas para realizar a identificação dos elementos relevantes dentro de um texto transcrito da fala natural.

1.1 Motivação

Atualmente quando uma situação de emergência criminal ocorre as pessoas ligam para os centros de emergência, onde podem falar com um operador da polícia que deve avaliar a situação e alocar recursos de acordo com a necessidade para resolver o problema apresentado.

Porém, acontecem alguns problemas que podem congestionar ou atrapalhar este serviço. Muitas pessoas ligam para essas centrais de atendimento de emergência em busca de informações como CEP de ruas ou outras informações, o que congestiona o serviço sem necessidade, já que esta informação não pode ser obtida por este meio. Outro problema enfrentado durante a tomada de decisão é a exigência de conhecimento do operador sobre o local onde está acontecendo a emergência e o conhecimento de diversos fatores ambientes e históricos que podem influenciar na situação atual e, conseqüentemente, na ação que deve ser tomada.

Este trabalho tem o intuito de auxiliar na filtragem destas informações e proporcionar uma base melhor para a geração de SAW, que irá apresentar novas informações, em tempo real, para um operador de sistema de análise de situações de emergência. Desta forma a tomada de decisão pode ter uma melhor sustentação informacional.

Com uma tomada de decisão melhor, menos recursos são gastos desnecessariamente e mais recursos podem ser usados em situações prioritárias. A melhor gerência de recursos podem resultar em menos perdas a vida e de bens, beneficiando a população com segurança e qualidade no serviço prestado.

1.2 Problema de Pesquisa

Com desenvolvimento deste trabalho foram exploradas técnicas para realizar o entendimento de fala natural (NLU), com o objetivo de prover subsídio para a geração de SAW de elhor qualidade.

Além das dificuldades encontradas dentro do domínio do NLU os dados utilizados para realizar as análises neste projeto são produtos gerados a partir de denúncias realizadas por humanos a serviços de emergência. A qualidade do dado obtido pode ser afetada negativamente por diversos fatores, que vão desde emocional da pessoa que esta relatando algo até eventos externos a pessoa que ocorram durante a obtenção dos dados.

Dados provenientes de humanos (HUMINT) são todos os dados que são gerados diretamente por humanos. Apesar de terem uma alta disponibilidade costuma ter uma baixa confiabilidade. Cada tipo de fonte de dados HUMINT apresenta particulari-

dades que precisam ser consideradas e tratadas de maneiras distintas.

Informações provenientes da fala natural tendem a apresentar problemas dentro da construção das frases. Durante a construção de falas eventos externos podem interferir o que faz com que o humano tenda a trocar entre frases subitamente, criar palavras, omitir palavras ou simplesmente parar no meio de uma frase.

Considerando a outra possível fonte que pode ser utilizada, as redes sociais, estas também relatam eventos ocorridos na vida real acabam resultando em textos que podem conter informações úteis. Os dados obtidos através de redes sociais costumam sofrer com erros sintáticos e gramaticais graves, o que pode gerar palavras desconhecidas e causar problemas durante as análises. O uso de gírias em redes sociais também é algo comum, e gera os mesmo problemas encontrados em palavras escritas de maneira incorreta. Para que se possa utilizar dados obtidos em redes sociais é necessário a realização de um pré-processamento para padronizar e corrigir a maior quantidade de problemas conhecidos.

O processamento da linguagem natural é um campo da computação que envolve diversos tipos de análises que podem ser realizadas em cima da fala natural de um humano, que vão desde definir a emoção de uma pessoa utilizando o tom de voz até realmente entender o que ela quer dizer com cada frase. Para a realização da tarefa de identificar elementos relevantes dentro de uma denúncia é necessário a realização do entendimento da fala natural (NLU).

A realização de NLU é uma tarefa muito complexa e que exige diversas análises. Como geralmente os dados disponíveis para esta tarefa são puramente textos transcritos, a identificação do sentido de cada palavra e frase é uma tarefa muito complexa, uma vez que não se tem o contexto da situação onde a fala foi feita.

Além da falta de contexto, outros pontos que podem atrapalhar durante a análise são: palavras desconhecidas, palavras com duplo sentido, ironia, falta de palavras, mudanças repentinas no texto, falta de pontuação, falta de tom de fala e grande complexidade das combinações para alcançar significados. Combinações de palavras em ordens distintas geram significados diferentes e muitas vezes até na mesma ordem podem ter o significado diferente.

Levar em consideração todos esses fatores é algo muito difícil e complexo. Este trabalho apresenta, em um domínio fechado, uma aproximação para a realização das análises com o menor impacto encontrado.

1.3 Justificativa

Quando uma ligação é feita para o serviço de emergência o operador que faz o atendimento dessa ligação tem como função identificar o que está acontecendo em uma determinada situação e decidir qual o melhor curso de ações a ser tomada para resolver o problema apresentado. Para poder decidir o que deve ser feito, o operador precisa entender quais elementos estão envolvidos na situação e qual a relação entre eles, ou seja, atingir um certo nível de SAW. Porém, muitos operadores não conhecem todos os problemas e todas as evoluções de cada um dos problemas que podem surgir, quais locais estão envolvidos em uma situação e qual o histórico do local ou dos elementos envolvidos e saber estas informações é algo muito importante para a tomada de decisão.

Possuir conhecimento histórico dos elementos envolvidos e possuir conhecimento geográfico do local onde está acontecendo a situação são fatores importantes para a geração de SAW para a tomada de decisão. Sem estes conhecimentos o mais completo possíveis, o operador pode tomar a decisão se baseado em dados que não representam a realidade de uma situação e com isso, gerar danos a vida e ao patrimônio.

Com o desenvolvimento deste trabalho, será possível melhorar o tempo de obtenção das informações usadas para gerar e dar manutenção a SAW, utilizando-se de denúncias feitas em fontes distintas. Com a melhora dos dados utilizados o operador pode melhorar a qualidade da decisão que será feita e assim poder resolver o problema de maneira mais eficiente.

1.4 Objetivo Geral

O desenvolvimento deste tem como objetivo geral propor um processo que consiste no uso de técnicas para a análise de dados HUMINT em busca de informações relevantes para o domínio escolhido. Com a identificação destas informações é possível gerar ou realizar a manutenção de SAW e auxiliar a tomada de decisão de um operador.

1.5 Objetivos Específicos

O trabalho desenvolvido e apresentado tem como objetivo específico procurar por elementos relevantes para o domínio de gerenciamento de emergência criminal realizando o entendimento de fala natural, para gerar subsídio de qualidade para SAW.

Para que o objetivo acima possa ser concluído, técnicas existentes e o desen-

volvimento de novas técnicas que sejam capazes de auxiliar durante a realização do entendimento de fala natural ou dados obtidos através de redes sociais. As técnicas utilizadas podem ser classificadas em sintáticas, gramaticais e semânticas e foram classificadas utilizando rede neural artificial com o intuito de facilitar o uso e melhorar os resultados.

1.6 Metodologia

O desenvolvimento deste trabalho conduzido de maneira exploratória e dirigido por um estudo de caso, mais especificamente uma situação de crime, onde é necessário identificar diversos elementos. Alguns dos elementos necessários para serem identificados são: criminoso, vítima, objeto roubado, arma utilizada e local. Para atingir os objetivos, as seguintes atividades foram desenvolvidas:

1. Estudo sobre o funcionamento do domínio de gerenciamento de emergências atualmente. Abordagem de técnicas de avaliação de situação e tomada de decisão. Identificação de elementos relevantes e suas relações para a tomada de decisão;
2. Estudo sobre o conceito de consciência situacional e relações existentes com gerenciamento de emergências. Influência da qualidade de dados utilizados para gerar consciência situacional e o entendimento da situação;
3. Estudo sobre técnicas para realização de análise de texto e busca de informações dentro de dados HUMINT;
4. Análise de técnicas de machine learning e estudo sobre a implementação destas técnicas nas análises de NLP realizadas neste trabalho; Implementação de rede neural artificial para a realização da análise sintática e gramatical;
5. Estudo sobre o estado da arte de gerenciamento de emergências, consciência situacional e técnicas envolvidas. Pesquisa sobre a aplicabilidade do entendimento de fala natural para auxiliar na tomada de decisão;
6. Pesquisa sobre a viabilidade da utilização de ontologia para adicionar semântica aos dados. Implementação e avaliação em conjunto com as análises que já utilizava dados sintáticos e gramaticais;
7. Avaliação dos resultados obtidos e comparação entre técnicas e resultados;

1.7 Organização

Durante o capítulo 1 foi apresentado a introdução deste trabalho. Na introdução está presente a motivação, problema de pesquisa, justificativa, objetivos e metodologia.

Na seção sobre a motivação foi discutido o porquê esse trabalho é importante e como ele pode contribuir para a sociedade e o meio científico. Durante o problema de pesquisa foi apresentado alguns dos problemas enfrentados durante o desenvolvimento. Durante a justificativa foi explicado o que esse trabalho pode fazer para melhorar o problema apresentado. Nas seções de objetivos e metodologia é discutido o que esse trabalho propõe a ser resolvido e os meios que ele utilizará para resolver este problema.

O capítulo 2 tem como objetivo apresentar a fundamentação teórica necessária para o entendimento deste trabalho. Os conceitos e técnicas das ferramentas utilizadas são explicados neste capítulo.

O capítulo 3 apresenta um estudo feito sobre trabalhos relacionados de alguma forma com este trabalho, podendo ter 1 ou mais pontos em comum. Comparações e análises são apresentadas.

Um passo a passo mostrando qual caminho e ordem deve ser seguida para a implementação deste trabalho. O processo de desenvolvimento é apresentado durante este capítulo com o objetivo de possibilitar o leitor de recriar os experimentos e resultados obtidos.

Os resultados gerados durante este trabalho estão presentes no capítulo 6 de forma detalhada, onde é possível comparar os resultados obtidos em diferentes momentos e com diferentes tipos de processos apresentados.

As conclusões obtidas com as implementações e resultados é discutida durante o capítulo 6, onde também é apresentado um diagrama para auxiliar na escolha da melhor técnica de análise.

Capítulo 2

CONSCIÊNCIA SITUACIONAL E FERRAMENTAS

Neste capítulo será apresentado o conceito, funcionamento e aplicação sobre consciência situacional e quais foram os modelos de apoio utilizados para o seu funcionamento.

2.1 Consciência Situacional

Consciência Situacional, ou Situation Awareness (SAW), é um processo cognitivo que contém o estado de uma situação e seus elementos em um determinado momento e sua possível evolução.

Utilizando técnicas para a criação e manutenção de SAW é possível, uma vez que a situação já esteja classificada, através do estado contido pelo SAW, compreender o que é e qual a função de cada elemento dentro da situação.

A criação, manutenção e entendimento de SAW é extremamente importante para todo tipo de processo onde para a tomada de decisão é necessário entender o que os elementos dentro de um cenário estão fazendo e como eles interagem entre si. Saber o estado atual de cada elemento e suas relações com outros elementos para a tomada de decisão é algo fundamental, uma vez que a tomada de decisão sem o conhecimento do ambiente onde a ação será realizada pode levar a cenários desastrosos.

2.2 Conceitos

Diversas definições foram apresentadas para a representação do significado de SAW. A definição apresentada por Endsley (1995) descreve SAW como "Saber

o que está acontecendo". Através desta definição é possível entender o que SAW faz: ter entendimento dos elementos dentro de uma situação. Assim é possível que durante o processo de tomada de decisão os elementos de um cenário possam ser consultados para saber qual a função deles, nesse cenário. De fato, SAW é saber o que está acontecendo em um determinado espaço de tempo e espaço, entender o que cada elemento influencia dentro de uma determinada situação e quais são as relações entre eles.

A utilização de SAW é utilizada constantemente em diversos cenários distintos, mesmo que em alguns a sua existência ocorra apesar de ser desconhecida. O uso de SAW acontece tanto por humanos quanto por máquinas. E isso pode ser observado em cenários para tomada de decisões médicas, onde as condições de um paciente podem indicar alguma doença e definir o tratamento, ou até mesmo para a tomada de decisão de planos de voos em aviões, onde o combustível disponível pode significar o encurtamento de uma viagem.

O entendimento humano do que acontece ao seu redor e como os elementos interagem entre si, apesar de não ser perfeito, é um tipo de SAW. Entender o que acontece ao redor do seu carro enquanto se dirige, saber quantas cadeiras ou janelas tem na sala em que se esta e qual a disposição delas é SAW. Além do entendimento humano, sistemas computacionais podem conseguir atingir SAW através de sensores que monitoram determinado ambiente em busca de informações.

Existem diversos sistemas computacionais atualmente que se aproveitam de SAW, como, por exemplo no setor médico ou de aviação. O uso de SAW em sistemas de diversos setores, como os apresentados anteriormente, auxilia de diversas maneiras as funcionalidades do sistema. Através do entendimento da situação atual e de sua possível evolução é possível se preparar e decidir curso de ações para a resolução de determinados problemas.

No meio de aviação o uso de SAW pode ser visto durante o voo de aeronaves, onde através de diversos sensores é possível saber a localização, velocidade, altitude e objetos próximos a aeronave. Isso é extremamente importante, pois assim se evita colisões, define cursos, diminui custos e outras funcionalidades. O mesmo tipo de preparo a partir de informações pode ser visto no meio médico. Tendo o SAW de um paciente é possível determinar o que tem de errado com ele, quais doenças podem estar presentes, como trata-lo, quais as opções possíveis e qual o melhor caminho a ser seguido.

Em um processo de tomada de decisão onde a ausência de SAW para a tomada de decisão é feita, o tomador de decisão precisa decidir o que será feito as cegas, isto é, ele decide que curso deve ser tomado sem saber o que acontece no ambiente onde a ação será executada. Seria como a torre de controle de tráfego aé-

reio informar para um avião que ele pode pousar sem consultar se a pista esta livre ou se outros aviões não vão tentar pousar ao mesmo tempo.

A tomada de decisão baseado em um SAW de má qualidade, que pode ser incompleto, desatualizado ou até mesmo inconsistente, pode resultar em uma tomada de decisão igualmente ruim a tomada de decisão sem SAW. A tomada de decisão com informações de SAW inconsistente, por exemplo pode levar um controlador de trafego aéreo fala para um avião pousar achando que a pista está livre, mas ela tem um avião tentando realizar a decolagem. A tomada de decisão com SAW desatualizado pode levar o controlador a falar para um avião que a pista esta livre para pouco, porem outro avião foi autorizado a realizar a decolagem momentos atrás.

2.3 Erros de SAW

A qualidade de SAW é diretamente dependente dos atributos contidos dentro daquele SAW e o quanto realmente representam a realidade. SAW com elementos e atributos discrepantes dos reais, ou seja, com erros, dentro de uma cena, geram SAW de má qualidade.

Quando SAW de má qualidade é consultado, as informações obtidas não podem ser usadas para a tomada de decisão em nenhum tipo de setor, tendo em vista que a tomada de decisão baseado em SAW de má qualidade pode resultar em tomada de decisão errada já que a situação representada não irá representar corretamente a situação real.

A tomada de decisão muitas vezes é baseada nas informações disponíveis em determinado momento sobre determinada situação, sendo assim o uso de SAW ajuda a interpretar a situação para decidir o curso de ação. O uso de SAW de baixa qualidade dentro do domínio de aviação, por exemplo, pode ser algo desastroso. Se uma aeronave tomar a decisão de reduzir a sua altitude baseando-se em informações erradas ela pode tentar reduzir a altitude em cima de prédios, montanhas ou simplesmente achar que o chão está centenas de metros abaixo dela e se chocar contra o chão. A ausência de elementos dentro do SAW também é um fator que pode gerar resultados desastrosos. No domínio da aviação, a ausência da posição de uma determinada aeronave pode significar inúmeras coisas e levar um controlador de tráfego aéreo, que pode autorizar ações que podem levar aviões a se chocarem.

O sucesso da tomada de decisão está ligado ao quanto o SAW utilizado é capaz de representar a situação no momento atual e sua evolução. A geração de problemas dentro do SAW podem ocorrer devido a alguns fatores.

2.3.1 Atualização de SAW

Conforme o tempo passa, a situação que foi analisada por SAW momentos antes evolui naturalmente, os elementos dentro da situação mudam e tendem a interagir de maneira diferente com os elementos que interagiam anteriormente e passa a ter novas interações. Se o SAW não evoluir juntamente com a situação ele passa a ficar desatualizado e apresentar diversos problemas, se transformando em SAW de má qualidade rapidamente a medida que se afasta da representar a situação atual.

A atualização de SAW consiste em atualizar as informações utilizadas para que o modelo mental gerado a partir destas informações continue representando corretamente as informações presentes na situação do cenário. A frequência dessas atualizações varia de cenário para cenário. As consequências da falta de atualização pode ter um alto custo para situações de emergência, por isso é de grande importância manter os dados utilizados para a geração de SAW sempre o mais próximo possíveis da realidade da situação.

O problema gerado pela evolução natural da situação pode ser resolvido através da manutenção de SAW, que irá realizar uma atualização dos dados contidos dentro do SAW para que não fiquem antigos e continuem a representar a situação atual corretamente.

2.4 Obtenção de SAW

A obtenção de SAW, ou seja, a geração do modelo mental a partir das informações disponíveis, permite que se entenda o contexto de uma situação, quais são os elementos envolvidos e como eles interagem entre si e prever como essa situação pode evoluir a partir dessas informações. Com SAW disponível a tomada de decisão pode ser feita baseada em informações consistentes e que realmente representam o estado atual da situação.

Como já constatado por outras pesquisas, SAW está presente em diversas situações onde a tomada de decisão se mostra como um fator decisivo ou objetivo final. A criação de um modelo mental capaz de conter a situação atual de um cenário ajuda durante a tomada de decisão, já que é possível entender com maior clareza o que está acontecendo na situação.

"... bombeiros, certas unidades de polícia, e o comando militar depende de SAW para tomarem suas decisões. Eles devem verificar recursos críticos em uma grande variedade de situações para determinar o melhor curso de ação. SAW imprecisa ou incompleta nesses ambientes podem levar a

desastrosa perda de vida, como no caso do U.S.S Vincennes.", (ENDSLEY, 1995)

A imagem 2.1, proposta por Endsley (1995), apresenta um resumo do funcionamento de SAW. Através dela pode ser percebido os diferentes níveis de SAW, a constante evolução e como cada decisão tomada influencia no estado atual de uma situação e em sua evolução. Como a cada momento um feedback é gerado as ações tomadas dentro da situação pode gerar novas situações, com características e elementos distintos da analisada inicialmente.

Com a geração de novas situações, criação de novos elementos e alteração constantes dos dados a serem gerados, a manutenção precisa a cada momento e a cada ação tomada dentro de uma situação é extremamente importante, já que, na falta da manutenção desses dados a tomada de decisão pode ser feita utilizando dados defasados.

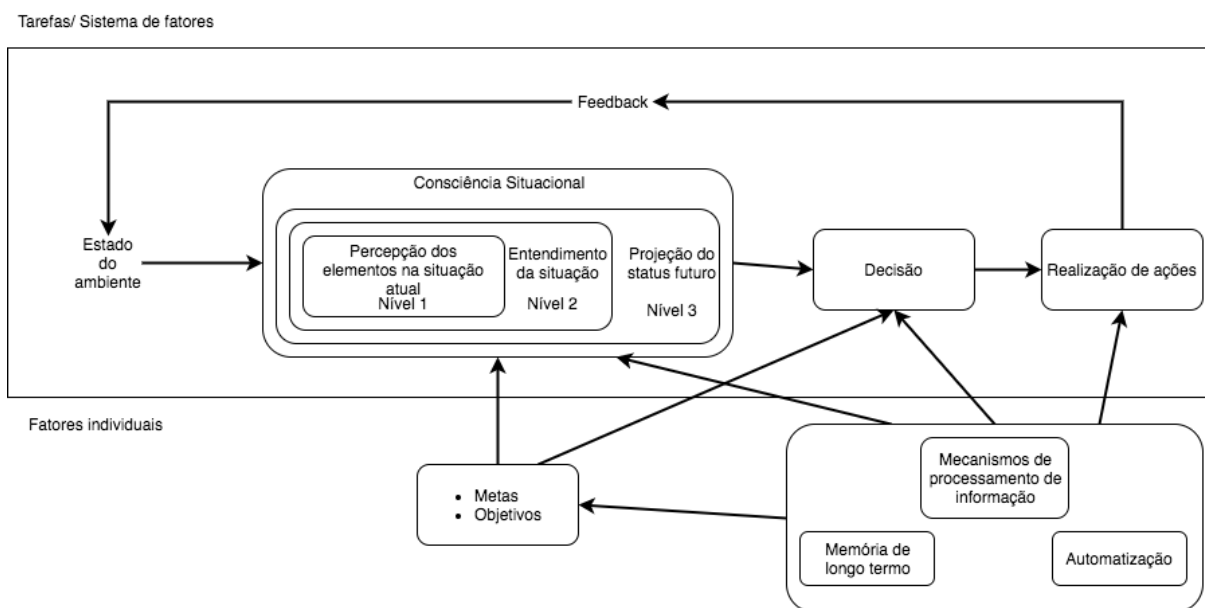


Figura 2.1 – Modelo de Consciência Situacional (Adaptado de Endsley, 1988).

É possível perceber a detecção dos elementos dentro de uma situação e como todo o fluxo de SAW interage com o ambiente para evoluir e se adaptar ao ambiente que está sendo representado. Para que se consiga utilizar de todo o potencial de SAW é preciso que a informação obtida consiga alcançar os 3 níveis.

O primeiro nível se constitui pela percepção dos elementos na situação atual que se procura obter o SAW, isso é, identificar cada elemento relevante e entender seu status dentro da situação atual. No domínio do estudo deste artigo um dos elementos

relevantes para se obter as informações é o criminoso e o que acontece com eles neste momento de tempo.

Dentro do primeiro nível os elementos são identificados dentro da situação através dos sensores e armazenados. Um avião, por exemplo, precisa antes de saber o que fazer identificar quais elementos tem para basear suas próximas ações, então aqui ele identifica suas informações como velocidade, altitude, direção e informações do ambiente ao seu redor, como relevo e clima, diferentemente do meio médico, que através das informações identificadas nessa etapa seriam o paciente, os ferimentos e quais instrumentos estão disponíveis na hora.

O segundo nível de SAW é formado pela atribuição de significados as informações obtidas no nível 1 e entender o que as ligações entre os elementos identificados significam. Nesta etapa é preciso entender como os elementos estão se relacionando e o que significa cada elemento. Um avião em voo precisa entender o que é uma montanha em determinada posição e qual a sua relação com outros elementos do ambiente como o próprio avião. No meio médico aqui o SAW identifica a informação de que um determinado ferimento no paciente esta causando determinadas reações. Os elementos identificados passam a ter relações e sentido nessa parte.

O terceiro nível é a projeção da situação no futuro através das informações reveladas pelos níveis 1 e 2. Esse é a parte de mais alto nível de SAW. Através dos itens identificados nos passos 1 e 2 é possível entender com completude o que está acontecendo em determinada situação em um espaço de tempo. Para que isso passe a ter mais valor é preciso através das relações e dos elementos identificados prever a possível evolução de uma determinada situação para que quando ela ocorra as ações adequadas para aquilo já tenham sido calculadas e possam ser apenas aplicadas.

Casos de funcionamento do terceiro passo é entender que o avião voando em determinada direção irá se chocar com o relevo. No meio médico é possível dizer nesta etapa que o ferimento do paciente se não for resolvido logo o levará a ter mais danos.

Os requisitos para a obtenção de SAW variam de cenário para cenário, já que as informações necessárias para se conseguir alcançar SAW variam em cada um deles. A obtenção e manutenção de SAW está ligada ao domínio da aplicação, geralmente quanto maior e mais dinâmico o domínio, mais difícil e complexo é a obtenção e manutenção destes dados.

Para a obtenção e manutenção de SAW para o domínio médico inclui saber a condição de um paciente, sabendo dizer seu estado atual e para onde o status dele ira progredir caso a situação permaneça de uma determinada maneira. Entretanto, para o domínio aviação alguns dos dados relevantes são: altitude, velocidade, objetos

próximos, inclinação e assim por diante.

Para o domínio deste trabalho, uma das fontes que está mais disponível é a denúncia realizada para os serviços de emergência. Para a obtenção destes dados, ferramentas que realizam a transcrição da fala natural foram utilizadas, bem como técnicas de NLP e NLU para a identificação dos elementos relevantes dentro das denúncias.

Através dos dados identificados do domínio específico é possível montar o SAW de um avião em voo e dizer que existem relevos perigosamente perto. No domínio médico é possível entender que um paciente está mal devido a um ferimento e o que a evolução desse ferimento irá causar ao paciente. Como cada domínio tem seus dados específicos de se identificar, dificilmente o SAW utilizada para analisar determinado domínio conseguirá contemplar um domínio diferente daquela.

No domínio de tomada de gerenciamento de emergências ter uma boa SAW é imprescindível para a tomada de decisão. Neste domínio a evolução rápida dos elementos precisa ser acompanhando o melhor possível. Uma tomada de decisão dentro deste domínio que esteja desatualizada pode resultar em prejuízos a vida ou patrimônio. Além de manter o estado atual sempre atualizado é preciso manter a precisão de cada elemento e suas relações o mais fiel possível a realidade, pois, elementos discrepantes também podem colocar a tomada de decisão em risco.

2.4.1 Manutenção de SAW

A manutenção de SAW se consiste pela atualização dos valores utilizados para se gerar SAW, isso deve acontecer de acordo com a evolução natural da situação. Em ambientes dinâmicos, onde os dados a serem analisados estão em constante evolução, essa análise se mostra particularmente mais difícil. Durante o tempo utilizado para a obtenção de SAW o ambiente pode já ter evoluído e agora os valores obtidos estão desatualizados e tomada de decisões baseados em valores desatualizados podem resultar em tomadas de decisões erradas.

A manutenção de SAW é de extrema importância para a tomada de decisão em qualquer domínio, já que o ambiente evolui e altera a situação e o estado. Com a evolução da situação o estado contido dentro do SAW fica desatualizado e volta a apresentar os problemas já discutidos.

A solução para manter SAW livre de discrepâncias da correta representação da realidade é atualizar as informações que são usadas para a geração de SAW a medida do necessário para o domínio e realizar a manutenção de SAW utilizando estas novas informações.

Como apresentado na imagem 2.1 após a representação da situação o ambi-

ente continua a evoluir e precisa ser reavaliado para que o SAW volte a representar novamente a situação real daquele momento. Por isso para que SAW seja válido ele precisa estar em constante atualização com o ambiente.

Dentro do domínio deste projeto para a realização da manutenção de SAW são as denúncias feitas ao serviço de emergência criminal. Para isso as denúncias precisam ser transcritas da fala natural e o conteúdo dos textos transcritos deve ser analisado em busca das informações que serão usadas como subsídio para a geração de SAW.

2.5 Objetivos de SAW

De acordo com os trabalhos feitos por Botega (2016) dentro do domínio de gerenciamento de emergência SAW se mostra muito importante para operações que envolvem humanos. Cita ainda a importância de não só caracterizar elementos, relações e ações, mas também prover subsídio para prever eventos futuros que podem causar qualquer tipo de prejuízo. Outro ponto destacado por Botega (2016) são os efeitos negativos sobre a tomada de decisão quando SAW está limitada, podendo levar a tomada de decisões errôneas para a segurança.

SAW tem a capacidade de manter um estado mental que representa uma situação em determinado momento, contemplando seus elementos e a forma como interação entre si. Além disso, SAW consegue dizer quais são as possíveis evoluções de uma situação. Através destes dados é possível se obter subsídio para a tomada de decisão envolvendo análises mais profundas de uma determinada situação a fim de evitar o erro, que em domínios de emergência pode ter prejuízos graves.

Com o uso de ferramentas para adicionar semântica aos dados, trabalhar com SAW é facilitado. Armazenar dentro do modelo mental os dados com semântica é algo que facilita quando é necessário realizar alguma consulta para prover subsídio para a tomada de decisão.

2.6 Representações de Informações para SAW

Ontologia é uma maneira de manter dados estruturados de uma forma que a estrutura atribui significado para o que está sendo guardado em cada parte, de forma que o computador consiga entender o que cada informação significa sem precisar de uma entrada de informações extras de um humano para isso.

De acordo com Botega (2016) com a utilização de ontologia é possível definir um domínio com semântica agregada. No trabalho ainda é discutido que como resultado da utilização de ontologias possibilita definições dos dados e que as rela-

ções entre eles ocorram de modo explícito. Através da utilização dessa ferramenta é possível que as informações sejam processadas por sistemas computacionais.

Os trabalhos desenvolvidos por Scharm et al. (2016) descrevem a ontologia como uma ferramenta para prover significado para o dado, a informação que pode então ser submetida ao processamento de algoritmos. Informações organizadas de maneira que possuem semântica é um grande atrativo, pois permite que o computador compreenda os elementos e suas relações dentro de um contexto. Além disso, maneiras de realizar consultas e armazenar informações com semânticas estão disponíveis e podem ser encontradas facilmente.

Com a ontologia alimentada de informações, as palavras dentro dela tem um significado e o computador consegue distinguir quando palavras escritas da mesma maneira tem significados diferentes. A ontologia é capaz de dizer se a palavra "banco" dentro de um contexto significa um banco de sentar, local onde se guarda o dinheiro, armazenamento lógico de dados ou até elevação no fundo de um corpo de água.

Além de ter conhecimento do significado das informações dentro de uma ontologia, é possível conhecer como determinado elemento pode se relacionar com outros elementos e quais relacionamentos estão acontecendo em um determinado momento. É possível saber que o elemento que contém as informações sobre uma vítima tem a relação de 'foi_roubado' pelo elemento que guarda as informações de um criminoso.

Para que a ontologia tenha a capacidade de realizar a distinção entre elementos e entender suas relações primeiro ela precisa ser ensinada. Antes do uso da ontologia dentro de uma aplicação ela precisa ser estudada cuidadosamente e a partir de casos reais e exemplos ela passa a ser construída. A construção da ontologia é o momento onde os especialistas no tema definem cada um dos possíveis objetos e a relação que ele pode ter com os outros elementos dentro de uma campo de atuação.

O tema abordado por Botega (2016) apresenta uma metodologia que se constitui em 7 passos definidos por Noy e McGuinness (2001). Através desta metodologia é possível alcançar níveis melhores para a ontologia de maneira mais objetiva.

Os passos definidos por Noy e McGuinness (2001) se constituem das seguintes etapas:

1. Determinar o Domínio e o Escopo da Ontologia
2. Reutilizar Ontologias Existentes;
3. Levantar termos importantes;
4. Definir classes e sua hierarquia
5. Definir propriedades das classes

6. Restrições das Propriedades

7. Criação de instâncias

As técnicas apresentadas neste trabalho tem como produto um conjunto de informações já salvas dentro de uma ontologia. A ontologia criada para a utilização dentro deste trabalho consiste em um conjunto de elementos e suas relações preparados para representar situações temporais no meio de gerenciamento de emergências militares. Com as informações salvas dentro da ontologias os dados estão prontos para serem enviados para as próximas partes do projeto continuarem com suas funções.

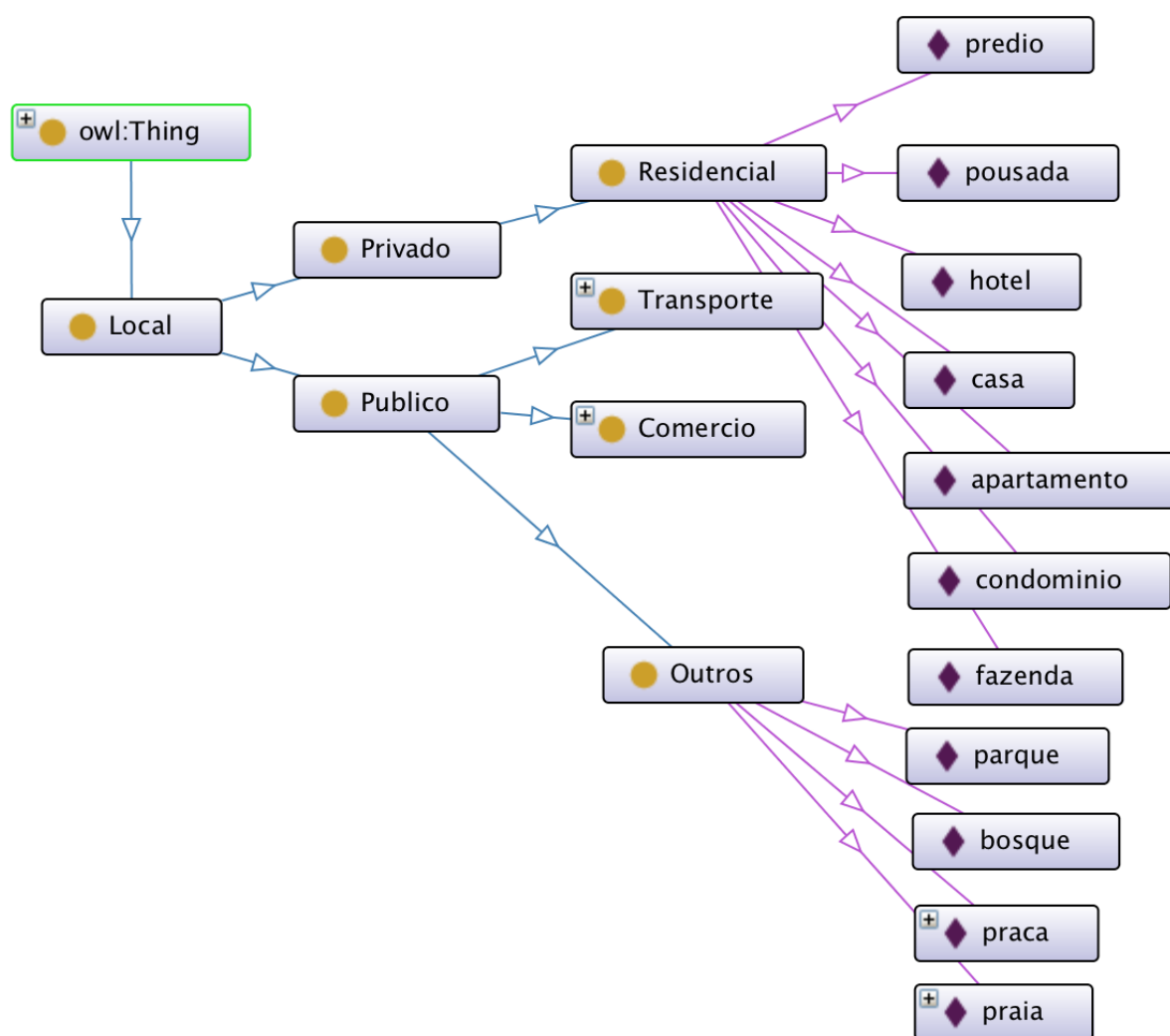


Figura 2.2 – Exemplo de ontologia

Dentro da ontologia apresentada na imagem 2.2 é possível ver como 'Residencial' é um elemento filho de 'Privado', isso significa que tudo que esta dentro de

'Residencial' automaticamente está dentro de "Privado". Cada elemento tem seus próprios atributos e relações. Através de cada seta que interliga dois elementos é uma relação. As ligações servem também para dizer quais as ações um elemento pode ter com outro elemento, além da ação de ser filho.

No final da árvore de exemplo utilizada é possível notar palavras de lugares que representam as instâncias da ontologia. As instâncias seriam os valores finais para cada galho da árvore, ou as folhas. Com a ontologia criada ela pode neste momento ser utilizada em conjunto com outras técnicas e ferramentas.

O uso de técnicas de SAW como explicado anteriormente possibilita que dados identificados a partir de uma situação, como os elementos dentro desta situação e as suas relações com outros. Porém para que seja possível o armazenamento computacional destes dados eles precisam ser salvos dentro do modelo de representação que a ontologia fornece.

A ontologia foi escolhida como técnica para a representação dessas informações pois tem a capacidade de apresentar os elementos, suas relações uns com os outros e não perder a semântica desses itens. A ontologia funciona como uma extensão de SAW e tem a função de prover técnicas que quando aplicadas conseguem representar as informações com semântica. É possível voltar para a ontologia em outros momentos do processamento e fazer requisições de informações de uma situação e a resposta será a informação já associada com a semântica dela.

A ontologia passa a ter as informações que foram identificadas dentro do SAW, já que ambas devem representar o mesmo conhecimento. Como a ontologia é um passo seguinte da geração de SAW ela passa a enfrentar os problemas discutidos na seção de problemas relacionados a SAW, precisando ser constantemente atualizada. As consequências para problemas não resolvidos dentro da ontologia podem levar aos mesmos problemas que SAW de má qualidade gera.

Com o uso de técnicas para a geração de ontologia é possível a utilização de inferência. O uso de inferências permite saber que um avião é um meio de transporte sem nunca essa informação ter sido dada explicitamente para a ontologia. Através de associações ela consegue chegar a essas conclusões.

Durante o desenvolvimento deste trabalho o uso da ontologia é feito em dois momentos distintos: durante a identificação de elementos relevantes e durante o armazenamento das informações identificadas. Durante a identificação dos elementos a ontologia é usada para consultar a semântica dos dados a serem analisados, o que auxilia na identificação dos elementos. Assim que os dados estão classificados eles são armazenados usando técnicas para manter a semântica das informações encontradas.

Para a realização das análises com semântica, a fim de encontrar os elementos relevantes é realizado o processamento de linguagem natural, juntamente com os dados gerados pela ontologia, para identificar os dados necessários para que ao fim do processo seja possível iniciar ou dar manutenção a SAW.

2.7 Processamento de linguagem natural

Processamento de linguagem natural, ou Natural Language Processing (NLP), é uma área da computação que tem como meio de trabalho a fala natural proveniente de humanos. Esses dados podem ser áudios que contém gravações de fala ou até mesmo textos.

"Processamento de linguagem natural (NLP) é uma área de pesquisa aplicada a explorar como computadores podem ser usados para entender e manipular textos de linguagem natural ou fala para fazer coisas úteis.", (CHOWDHURY, 2003)

Chowdhury (2003) Apresenta dentro de seu trabalho um tópico específico, "Some Theoretical Developments", sobre técnicas para a realização da análise no texto. Dentro desta seção são apresentadas diversas técnicas de diversos autores falando sobre algumas de suas características. O trabalho pode ser usado como referências para o estudo e implementação de outras técnicas não abordadas neste trabalho.

O uso de NLP está cada vez mais presente, uma vez que consegue criar uma interface mais natural para a entrada de dados em um computador. Essas entradas mais naturais podem ocorrer de diversas maneiras, indo desde uma pessoa falando e dando ordens para o computador até a interpretação de textos que foram escritos por humanos e não possuem nada além de texto para ser lido, algo que acontece muito comumente na internet atualmente.

Cada área de pesquisa dentro de NLP tem diferentes objetivos, algumas tratam de áudios, textos transcritos e até outros tipos de dados relacionados a fala. Dentro deste trabalho o foco é a obtenção de informações dentro de um texto transcrito proveniente da fala natural de um humano. Desta maneira, os dados de entrada que serão utilizados são apenas textos sem informações adicionais relacionadas ao que está escrito dentro de cada caso.

Devido ao domínio do trabalho os dados apesar de serem trabalhados como textos transcritos não chegam como texto. Os dados são ligações telefônicas que devem ser transcritos para serem utilizados. Para a realização dessa transcrição ferramentas específicas foram utilizadas, delegando assim a função de entender o áudio

e transforma-lo em texto. A ferramenta utilizada para realizar a transcrição é o: Google Speech To Text.

Uma vez com os dados transcritos, as análises podem ser iniciadas. Algumas das ferramentas que fizeram parte durante as análises terão suas funções descritas abaixo.

2.8 Ferramentas Utilizadas para NLP

Durante esta seção serão discutidas algumas das ferramentas que foram utilizadas durante o desenvolvimento deste trabalho para realizar o processamento de linguagem natural.

2.8.1 Google Speech To Text

Google Speech To Text é uma ferramenta disponibilizada pela empresa Google que consegue transcrever áudios de pessoas falando em texto em tempo real.

Uma das vantagens de se utilizar esta ferramenta é o acesso automático a todo o vocabulário que ela já possui, isso permite que a preocupação durante o desenvolvimento do trabalho se foque no desenvolvimento das técnicas pós transcrição do áudio e não na maneira de fazer essa transcrição e no que é precisa transcrever. Os modelos criados pela empresa também são aproveitados.

A necessidade de transcrição em tempo real vem atrelada a necessidade da velocidade de análise das informações para a tomada de decisão. Durante testes realizados para a escolha desta ferramenta foi constatado também uma boa precisão quando comparado com outras ferramentas do tipo para a transcrição de nomes de ruas, mesmo quando em outras línguas.

2.8.2 Cogroo

O Cogroo se trata de uma acoplável em um editor de texto como um *plugin*. Tem como função fazer análise gramatical do texto dentro deste editor e sugerir edições para os usuários.

Apesar de ter sido desenvolvido originalmente como um corretor para um editor de texto dentro de seu código fonte tem um parser, que é uma ferramenta para análise de texto, o qual pode ser usado para adicionar informações as palavras dentro de uma frase. Algumas das informações adicionadas são classe gramatical, ligação entre palavras e variações de uma palavra. O funcionamento do Cogroo pode ser observado no trabalho desenvolvido por Kinoshita, Salvador e Menezes (2005).

2.9 Natural Language Understanding (NLU)

Natural Language Understanding é uma sub-área de NLP que trabalha com a obtenção de significados de sentenças e palavras. O objetivo de NLU é replicar a capacidade humana de obter o significado desses dados.

"Quanto uma pessoa vê ou ouve uma sentença, ela faz completo uso de seu conhecimento e inteligência para entender. Isso não inclui apenas gramática, mas também o seu conhecimento sobre palavras, contexto da sentença, e mais importante, sua compreensão do assunto. Para modelar esse processo de entendimento de linguagem em um computador, nós precisamos de um programa que combine gramática, semântica, e raciocínio em um jeito íntimo, concentrando em suas interações.", (WINOGRAD, 1972)

Replicar a capacidade humana de realizar o entendimento pleno de uma sentença é algo de grande dificuldade, uma vez que requer a associação de diferentes habilidades humanas para se conseguir atribuir significados aos textos de entrada. Por isso diferentes técnicas precisam ser aplicadas a fim de realizar a classificação de maneira correta do texto.

2.10 Classificação de texto

Classificação de texto, ou Text Classification (TC), como descrita por Sebastiani (2002) é uma área de NLP e NLU que realiza a classificação de textos que foram gerados através de linguagem natural em categorias pré-definidas. Este tipo de trabalho tem ganhado cada vez mais relevância devido a crescente necessidade de se entender textos na internet, que não possuem nenhum tipo de metadado além do que foi digitado pensando que apenas pessoas iriam ler aquele texto.

O entendimento de dados provenientes de humanos (HUMINT) tem ganhado relevância frente ao mercado e pesquisa mundial devido ao advento de sistemas que interagem com humanos estarem cada vez mais presentes. Com a chegada de ferramentas para interação com o humano a interface dessa comunicação é um fator essencial, isto é, como o humano consegue interagir com os sistemas. A análise da fala humana e de textos é uma forma fácil e natural para que humanos consigam se comunicar com tais sistemas.

Apesar de ser natural e diminuir a curva de aprendizado para humanos, o entendimento de dados HUMINT sempre é um fator extremamente difícil de ser analisado. A inconsistência dos dados, incoerência, falta de informações são fatores que podem durante a análise destes dados serem interpretados erroneamente.

Sebastiani (2002) também explica a evolução das técnicas de leitura e entendimento destes textos e a seu recente alinhamento com técnicas de machine learning.

2.11 Obtenção de dados para SAW através de classificação de texto

A utilização da realização de NLP pode ser aplicada de diversas formas para a busca de diversas informações diferentes. Para a utilização de NLP com o objetivo de prover informações para a criação, manutenção e utilização de SAW os fatores que influenciam na qualidade de SAW vão influenciar nos métodos e objetivos do NLP utilizado.

Como SAW a qualidade de SAW é dependente da representação fiel de uma situação os dados analisados pelo NLP devem seguir o mesmo caminho, e assim, procurar ao máximo encontrar as informações necessárias para a geração de um SAW de boa qualidade. As informações procuradas para a criação de SAW de qualidade devem ser transferidas em informações que a NLP deve buscar. Todos os elementos devem ser informados a NLP para que ela se preocupe apenas em encontra-los.

É possível que o sensor que irá prover as informações para a realização da NLP traga muitas informações adicionais, além das informações que devem ser encontradas para a criação do SAW. A parte responsável por realizar o NLP pode se preocupar em encontrar os elementos necessários para a criação de SAW. Em um domínio de tomada de decisão aéreo por exemplo não é relevante para a SAW de um avião que houve um acidente de carro em uma rodovia.

2.12 Inteligência artificial

Inteligência artificial é um paradigma de programação em que o computador não é programado com formulas de como resolver determinado problemas, mas sim com técnicas que permitam o aprendizado de como chegar a uma determinada solução. Programas que tem algum tipo de inteligência artificial tem o uso de seu código dividido em dois momentos distintos. O momento de treino e o momento de aplicação. Durante o treino da inteligência o programa tem a capacidade de aprender com os dados que são fornecidos. Só após o treino o código pode ser usado para aplicação.

Durante a fase de treinamento dados são disponibilizados para serem analisados pelo programa, que tem como função utiliza-los para aprender a classifica-los de uma forma que quando dados parecidos com eles apareçam eles consigam generalizar o que foi aprendido e classificar esses dados novos.

Durante a fase de aplicação dados são disponibilizados para o programa que

precisa ser capaz de lê-los e dar uma saída baseado nos dados que aprendeu durante a fase de treinamento. Através de proximidade de generalização o programa deve ser capaz de dizer qual a saída adequada para os dados de entrada.

Inteligência artificial monta de maneira implícita modelos matemáticos capazes de resolver os dados de que são disponibilizados e gerar uma saída. Através das entradas de treinamento o computador consegue criar métodos para generalização, assim pode identificar dados que não foram avaliados anteriormente.

Através do treinamento os algoritmos de aprendizado de máquina criam modelos que quando determinado conjunto de entradas forem disponibilizados, através de diversas técnicas e funções matemáticas, as saídas devem ser aproximada das saídas ensinadas na fase de treinamento.

2.13 Redes Neurais Artificiais para Identificação de Elementos Relevantes

Redes neurais artificiais são recursos computacionais que passaram a existir após estudos sobre o funcionamento do cérebro e como ele funciona. Os principais recursos de uma rede neural artificial são os "Neurônios" e as ligações entre eles, que guardam o peso entre os neurônios.

A entrada de dados passa pelos neurônios e seu valor é multiplicado pelo peso da ligação. O resultado disso é somado no neurônio na outra ponta desta ligação e assim por diante. Quando todos os neurônios de uma camada já foram processados a rede procede para a próxima camada e assim por diante até chegar nas saídas. Uma rede pode ser dividida em 3 partes: Entrada de dados, Hidden Layer, Saída de dados.

A entrada de dados é por onde os valores para serem processados pela rede neural serão disponibilizados. Geralmente é constituída por uma camada onde todos os neurônios são entradas e pode ter entradas ilimitadas.

O hidden layer se constitui por X camadas com Y neurônios cada, onde os neurônios se comunicam entre si. Não é obrigatório a existência do hidden layer em uma rede, apesar de proporcionar a solução de problemas mais complexos. A saída da rede se constitui novamente de N neurônios e contem os resultados finais após o completo processamento da rede.

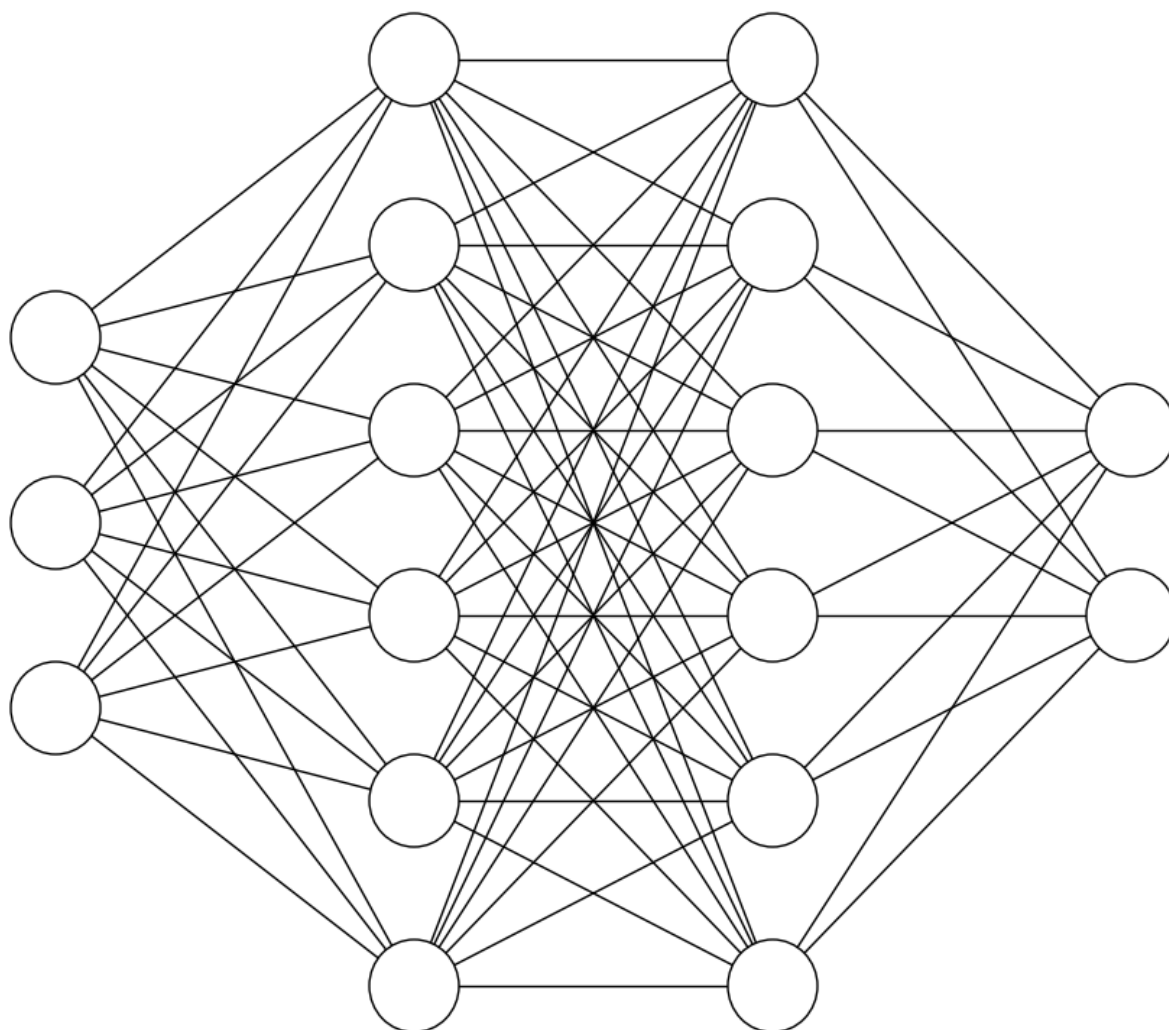


Figura 2.3 – Representação de uma rede neural com 3 entradas, 2 hidden layer de 6 neurônios cada e duas saídas

A imagem 2.3 apresenta uma rede neural onde podem ser identificado claramente as entradas do lado esquerdo da imagem, as saídas do lado direito e no meio os hidden layer.

As redes neurais artificiais passam pelas mesmas etapas de treinamento e aplicação que algoritmos de aprendizado de máquinas tem que passar antes de serem utilizadas. Existem 3 principais métodos de aprendizado de máquina para redes neurais: Aprendizado assistido, aprendizado não assistido, aprendizado por reforço.

O aprendizado assistido é definido por um aprendizado onde é disponibilizado a entrada desejada e a saída desejada para aquela entrada de dados. Diversos casos são disponibilizados e a rede neural deve aprender a montar padrões onde quando a entrada do exemplo aparecer a saída deve ser o mais próxima possível da disponibilizada na fase de treinamento.

O aprendizado não assistido a rede neural recebe apenas um conjunto de dados e nenhuma saída. A sua função neste método é aprender a montar padrões de agrupamento sozinha, assim quando chegar a fase de aplicação a rede checará de onde o dado de entrada está mais próximo e o classificará desta maneira.

O aprendizado por reforço consiste em um conjunto de dados é disponibilizado e para cada dado a rede classifica e apresenta uma saída, um aviso é dado para a rede para falar se a saída apresentada por ela foi boa. Se a saída for classificada como boa a rede reforça as ligações utilizadas para chegar nesse valor.

Apesar dos avanços tecnológicos permitirem o uso de redes neurais com mais facilidade, o seu tamanho, tamanho de entradas, hidden layers e saídas, ainda é, muitas vezes, limitado pelo poder computacional disponível. Durante a fase de treinamento o tempo usado é relativamente alto, estando diretamente relacionado com o tamanho da rede neural. Geralmente redes maiores tendem a demorar mais tempo para serem treinadas.

Apesar do tamanho das redes neurais não ser infinito isso não é necessariamente uma limitação delas. Geralmente redes relativamente pequenas conseguem resolver grandes problemas de maneira simples. Um único neurônio é capaz de resolver problemas linearmente separáveis. Quando o problema passa a ser não linearmente separável é necessário o uso de uma rede contendo mais que um único neurônio. Redes com 2 entradas e uma saída consegue resolver problemas como um XOR lógico. Redes com 2 entradas, 3 hidden layer de 20 neurônios cada e 3 neurônios de saída já é capaz de realizar pinturas com cores RGB complexas.

"Redes neurais artificiais (ANNs) é um dos mais comuns em aprendizado de máquina, ele simula a rede neural do corpo humano que consiste em muitas camadas de neurônios. [...] Um grupo de neurônios em uma rede neural tem funções separadas ao mesmo tempo. Em redes neurais existe o estado de aprendizado que consiste no ajuste dos pesos para obter a saída desejada e o estado de teste que a rede neural é testada para ver sua precisão no processo de detecção.", Al-hadidi, Alarabeyyat e Alhanahnah (2016)

Como citado por Chang e Tsai (2017) em seu trabalho com os avanços da tecnologia é possível atualmente treinar redes neurais artificiais em pouco tempo e sem depender de semanas para que isso seja feito, algo que alavancou o uso de soluções deste tipo.

Durante o desenvolvimento deste trabalho o uso de treinamento supervisionado foi escolhido devido a disponibilidade de dados classificados e a grande possibilidade

de entradas. O aprendizado supervisionado se constitui na disponibilização das entradas para a rede e a saída esperada para cada conjunto de entrada. Sempre que a saída esperada é diferente da obtida os valores das ligações entre os neurônios é alterada. Isso é feito através de teste entre as entradas disponibilizadas até que o erro da rede seja menor que um valor pré-definido.

A grande vantagem que a adoção de técnica como redes neurais artificiais apresentam é a sua capacidade de aprenderem a solucionar os problemas propostos utilizando modelos não lineares e a sua capacidade de generalização. Assim após a fase de treino é possível apresentar uma solução para um problema que nunca foi apresentado durante a fase de treino e obter a solução correta. Através da generalização uma rede neural consegue dizer quando a entrada disponibilizada se aproxima a algum dos treinos usados anteriormente.

2.13.1 Ferramentas em Suporte a Rede Neural Artificial

As ferramentas utilizadas para a criação da rede neural artificial utilizada durante o desenvolvimento deste projeto será descrita abaixo

2.13.1.1 Encog

Encog é um framework que auxilia na criação e manejo de ambientes que possuem aprendizado de máquina. Durante o uso do Encog as ferramentas utilizadas foram a criação de uma rede neural, que com poucas parâmetros é possível criar e começar a utilizar a rede. A linguagem de programação escolhida foi javascript devido a sua facilidade. Encog também possui suporte para Java, .NET e C/C++.

2.13.2 Redes neurais e processamento de linguagem natural

Como discutido anteriormente, modelos de redes neurais podem ser traduzidos como funções matemáticas capazes de solucionar problemas. Então, devido ao cenário enfrentado nesta pesquisa, os dados da situação analisada, que não tem nenhum tipo de informação matemática, precisa ser convertido para dados numéricos para que possam ser inseridas nas entradas das redes neurais.

O objetivo de se aliar redes neurais com análise de texto aparece quando a análise contempla um número muito grande de informações para que todas possam ser analisadas e implementadas manualmente, o que geraria milhares de trechos de código contendo diversos iterações e repetições para se analisar casos específicos, o que pode gerar diversos problemas. Uma maneira automatizada de se fazer isso poderia ganhar tempo e evitar problemas gerados por análises e implementações manuais de cada caso de análise do texto.

Com a utilização de uma técnica, é possível transformar os dados textuais em números que podem ser utilizados como entrada para as ANN. Esta técnica é responsável por determinar qual entrada corresponde a cada possível palavra e sempre que esta palavra estiver presente associar a entrada da ANN o valor 1.

Com a utilização de ANN para realizar a classificação dos dados foi possível eliminar as dezenas de iterações e loops que era utilizados para realizar a identificação de cada palavra ou conjunto de palavras, já que as redes neurais podem ser treinadas com exemplos antes de seu uso. Uma vez treinada, a rede neural pode realizar a classificação de maneira fácil e rápida dos dados. Isso significa que, o trabalho gasto estudando as possibilidades e criando a análise individual de cada uma pode, agora, ser gasto a procura de novos exemplos para melhorar os resultados da classificação da rede neural, que pode ser capaz de identificar mais coisas.

Como resultado após a fase de treinamento, a rede é capaz de receber um conjunto de palavras e dizer a qual determinado elemento ela se refere, desde que tenha sido treinada para realizar a identificação desse elemento alguma vez.

A ANN utilizada leva em consideração os dados sintáticos, gramaticais e semânticos das palavras e frases disponíveis para que, baseado nos treinos, seja capaz de classificar os dados em elementos.

Capítulo 3

PROCESSAMENTO E ENTENDIMENTO DE LINGUAGEM NATURAL NO GERENCIAMENTO DE EMERGÊNCIAS PARA OBTENÇÃO DE CONSCIÊNCIA SITUACIONAL

Neste capítulo será discutido o processo de desenvolvimento feito neste trabalho. As técnicas utilizadas serão discutidas, bem como sua utilização e funcionamento.

3.1 Trabalhos Correlatos

Durante a revisão bibliográfica foi constatado que existem diversos trabalhos já produzidos e que contribuíram, de forma direta ou indireta, para esta produção. Existe também uma grande variedade de trabalhos em que um ou dois temas discutidos neste trabalho são apresentados, porém, a quantidade de trabalhos já produzidos que discutem a união entre processamento de linguagem natural em situações de emergência utilizando rede neural, parsers e semântica, são extremamente escassos, por serem muito específicos. Abaixo seguem trabalhos relacionados com alguns dos temas apresentados neste.

No trabalho apresentado por Collobert e Weston (2008) foi discutido uma rede neural capaz de aprender por multitarefa diversas informações ao mesmo tempo. O objetivo desta rede neural desenvolvida é a realização de uma série de classificações que são feitas em textos provenientes da fala natural. Essas classificações podem ser sintáticas, gramaticais ou quanto à semântica, em algumas classificações. Foram feitos testes utilizando dados disponíveis em locais como Wikipédia para a realização

do treinamento, utilizando uma base de dados com milhões de palavras.

Os bons resultados obtidos no trabalho mostram que redes neurais co-evolucionais podem se beneficiar de diversos fatores, como a grande quantidade de dados disponíveis e os dados utilizados para a realização das análises. O trabalho apresenta também uma transformação de índices para vetores, que mapeia as palavras em vetores para poderem ser usadas como entrada para a rede neural.

Em situações de emergência, o trabalho desenvolvido por Verma et al. (2011) apresenta como foi possível realizar a extração de dados, necessária para a formação de SAW. Para a realização dessa tarefa foram utilizados dados inseridos em redes sociais. Após os dados serem obtidos do Twitter, eles puderam ser processados usando técnicas de processamento de linguagem natural para realizar a busca por informações.

O trabalho apresentado por Verma et al. (2011) também realça a quantidade de dados gerados que podem ser relevantes na rede social utilizada durante os eventos de emergência abordados e a dificuldade de realizar a busca automatizada pelas informações necessárias. Primeiramente, a seleção dos dados foi feita através de palavras-chave, que foram definidas para o domínio da aplicação. As técnicas de aprendizado de máquina Naïve Bayes e Maximum Entropy foram utilizadas por serem conhecidas pela boa realização da tarefa de classificação do texto e de reconhecimento de emoções. Como resultado disso, o trabalho conseguiu prever quando um dado terá informações relevantes para a geração de SAW e classificá-las.

Durante o desenvolvimento do trabalho de Matheus, Kokar e Baclawski (2003) foi proposta uma ontologia que representa uma situação de emergência. Além disso, outras variações foram propostas com o objetivo de escolher a que melhor se corresponde à necessidade existente. As variações surgem com o objetivo de melhorar alguns pontos, mas podem ter desvantagens quando as mudanças ocorrem. O ganho de facilidade, por exemplo, acarreta o armazenamento de informações redundantes.

As ontologias propostas são capazes de representar os objetos presentes em uma situação e as suas relações, bem como a evolução dos mesmos.

Com o estudo realizado sobre os trabalhos já realizados presentes na bibliografia ficou clara a falta de trabalhos que abordem todos os temas necessários para o desenvolvimento dessa pesquisa em um único trabalho, ou seja, trabalhos que de fato abordam o tema abordado por este trabalho. Foi constatado também que, mesmo tirando o domínio criminal, os temas restantes ainda sofrem com a falta de trabalho onde aparecem juntos. Considerar a realização da classificação semântica dos dados utilizando

3.2 Desenvolvimento

Durante o desenvolvimento deste trabalho a evolução ocorreu com 4 tipos de análises, onde em cada um deles foi desenvolvido e testado uma análise diferente e com informações diferentes. Com o desenvolvimento destes estágios, onde o estado seguinte é sempre uma evolução do anterior, foi possível determinar qual a melhor maneira de realizar a identificação de elementos dentro do cenário utilizado.

Como apresentado durante a metodologia, os estágios de desenvolvimentos podem ser separados em 3 momentos distintos

1. Pesquisa sobre áreas utilizadas, onde as tecnologias e técnicas existentes foram estudadas
2. Desenvolvimento e implementação das técnicas para a realização de elementos relevantes em textos transcritos a partir da fala natural.
3. Avaliação dos resultados gerais e definição da melhor técnica para este trabalho

Durante o estágio de pesquisa foram feitas pesquisas sobre todas as áreas que estão envolvidas no projeto. As áreas de gerenciamento de emergência, consciência situacional, processamento de linguagem natural e machine learning foram estudadas.

No estágio de desenvolvimento foram implementadas as técnicas desenvolvidas e técnicas encontradas na literatura, de forma separada e em conjunto. Durante a avaliação de resultados, os resultados obtidos após as implementações foram comparados e classificados para que fosse possível realizar a classificação e a identificação dos pontos positivos e negativos de cada um. Os tipos de processamento que foram utilizados durante as análises se constituem por: "Análise Sintática", "Análise Sintática e Gramatical", "Análise Sintática, Gramatical e Semântica", "Análise Sintática, Gramatical, Semântica com Rede Neural".

O objetivo de todas as técnicas utilizadas para realizar o entendimento de linguagem natural é o mesmo: identificar informações utilizando como dado de entrada um texto transcrito da fala natural. Com esse objetivo comum, diferentes técnicas foram desenvolvidas e evoluídas para se obter os melhores resultados possíveis.

Durante a realização de todas as análises feitas durante este trabalho, apresentadas nas seções a seguir, o fluxo seguido pelos dados é representado pela imagem 3.1. A primeira e a última caixa do desenho, "Obtenção de dados" e "Salvar Dados Gerados", são padrões para todas as análises. As seções a seguir também discutem as alternativas envolvidas com o bloco de "Análise" mencionado na figura abaixo.

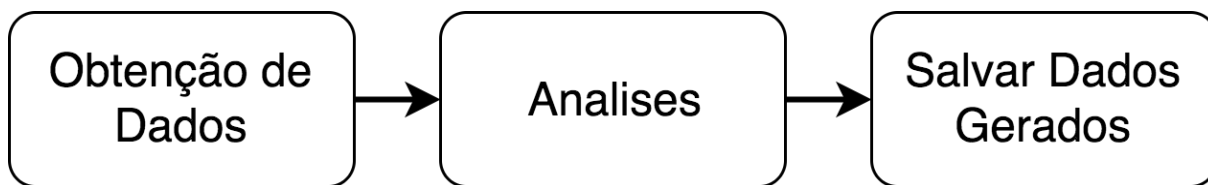


Figura 3.1 – Fluxo Geral das Análises

As análises discutidas terão os mesmos tipos de obtenção de dados e geram saídas parecidas, já que só diferem na forma de salvar o resultado gerado. Durante a obtenção dos dados é feita a aquisição dos dados por redes sociais ou a transcrição da fala natural para texto. A caixa correspondente a 'Salvar Dados Gerados' tem a função de salvar os resultados gerados pelas análises e avisar os módulos seguintes.

A lista abaixo apresenta algumas das denúncias que foram utilizadas durante o processamento. Esta lista é composta por denúncias feitas ao serviço de emergência criminal e textos obtidos em redes sociais. Os textos obtidos através das redes sociais presentes abaixo estão já pré-processados, para que possam ser analisados normalmente dentro do fluxo deste trabalho.

1. "um crime acaba de acontecer aqui na avenida domingos setti um motorista foi ameaçado e expulso do carro sem levar nada o ladrão fugiu em direção ao metrô klabin
2. "um cara foi assaltado na minha frente por alguém armado foi na domingos setti perto do restaurante don paladino"
3. "acabo de ver uma senhora sendo assaltada aqui na praça da sé o meliante levou sua bolsa"
4. "um cara de camisa azul ameaçou e roubou uma senhora perto da praça da sé ele tinha uma faca"
5. "acaba de acontecer um crime aqui na domingos setti um motorista foi ameaçado e mandaram ele sair do veículo sem levar nada o bandido fugiu em direção ao metrô klabin"
6. "acabei de presenciar dois homens roubar um mercedes preto aqui na domingos setti ameaçaram o senhor entraram no carro e fugiram em direção ao metro kabin"

3.3 Identificação de Elementos com Análise Sintática

Durante o primeiro estágio para realizar o processamento dos dados, foram utilizadas técnicas de análise sintática. Este tipo de processamento se constitui na análise da formação das frases e palavras, considerando a combinação de símbolos existentes, isto é, a análise sintática considera a ordem que as palavras são escritas, sem levar em consideração nenhum tipo de contexto, gramática ou semântica para isso.

Durante a análise sintática apenas as comparações mais simples eram feitas. Para determinar se algo estava dentro de uma frase apenas a busca por palavras associadas a este elemento era buscada e sempre que eram encontradas essa palavra era separada. Para se determinar as palavras que deveriam ser buscadas dentro da frase, uma pesquisa utilizando denúncias reais foi realizada e algumas das palavras chaves foram identificadas.

O fluxograma na imagem 3.2 mostra como a parte que realiza a análise sintática pode ser representada com apenas um bloco, já que realiza apenas um tipo de análise em apenas um momento. A entrada de dados e a saída dos resultados gerados são separados pela análise que faz toda sua análise de maneira síncrona.

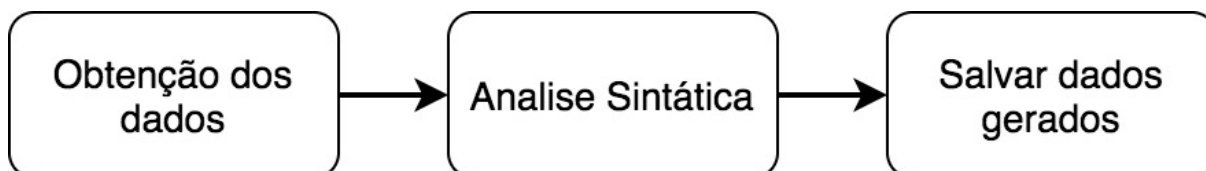


Figura 3.2 – Identificação de elementos utilizando apenas análise sintática

O pseudo-código apresenta como é feita a análise sintática, contendo pequenos exemplos das palavras que estão sendo buscadas dentro dele.

```
1 var carros_encontrados = []
2 var armas_encontradas = []
3
4 var carros_palavras = ['carro', 'mercedes', 'audi', 'bmw']
5 var armas_palavras = ['arma', 'faca', 'rifles', 'metralhadora']
6
7 para (var i=0; i<tamanho_frase; i++){
8
9     se (carros_palavras.contem(frase[i])){
10         carros_encontrados.adiciona(frases[i])
11     }
12     se (armas_palavras.contem(frase[i])){
13         armas_encontradas.adiciona(frases[i])
```

```
14     }  
15 }  
16  
17 escreve carros_encontrados  
18 escreve armas_encontradas
```

Listing 3.1 – Pseudo-Código de Análise Sintática

Este é o modo mais simples de se procurar informações dentro de frases, porém, pode conter diversas inconsistências. Como a busca é basicamente uma checagem se determinadas palavras existem dentro da frase sendo analisada, às vezes, inconsistências podem ser geradas, como quando a palavra “arma” é encontrada é marcado que se tem uma arma na denúncia, mas pode acontecer de que às duas palavras anteriores a arma seja “não tem” e assim o resultado deveria apontar o oposto disso.

Buscas mais complexas, como as buscas que realizam algumas checagens de relações entre palavras e proximidade de palavras, podem ser realizadas, porém a grande dificuldade, falta de semântica e problemas com a geração de cada caso específico tornam isso um trabalho inviável para frases que podem conter qualquer informação. O resultado deste tipo de análise deve ser definido dentro dos ‘SE’ quando uma palavra é encontrada. Geralmente a saída desta análise é um vetor contendo palavras encontradas.

3.4 Gramática para Apoiar a Busca de Elementos Relevantes

Como solução para melhorar as análises feitas o uso de informações gramaticais passa a ser feita neste estágio. A ideia por trás do uso de informações gramaticais é aumentar o nível de abstração durante a criação e utilização das análises, isto é, ao invés de se conhecer todas as palavras que seriam possíveis de se usar para descrever a cor de um objeto e criar casos para a identificação de cada uma delas, é possível agora considerar apenas a classe “adjetivo”.

A imagem 3.3 mostra que após a obtenção dos dados a classificação gramatical é feita e apenas após esta classificação estar completa, a análise, que acontece de maneira paralela e considera os dados sintáticos e gramaticais disponíveis, é realizada.

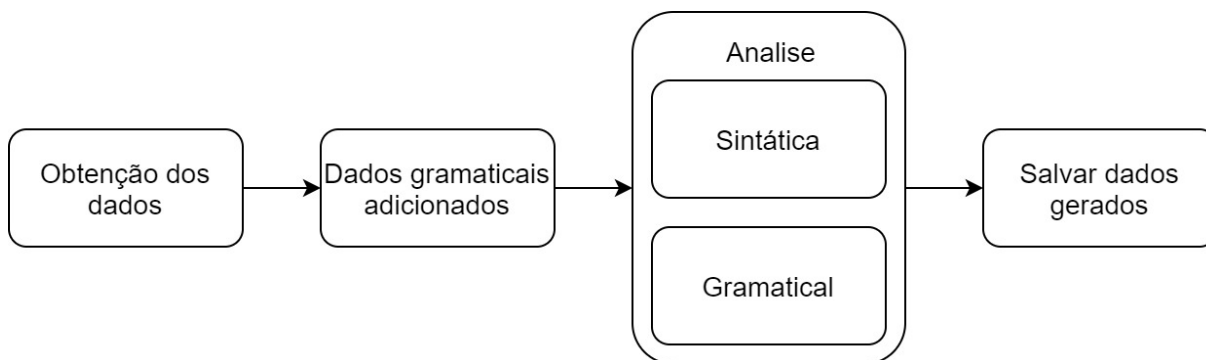


Figura 3.3 – identificação de elementos relevantes com análise sintática e gramatical

Com o uso desse tipo de informação a generalização do código foi facilitada, aumentando a velocidade e eficiência das análises realizadas. O exemplo abaixo mostra como fica a classificação gramatical da palavra “moto”.

```
1 {
2   "tokens ":[
3     {
4       "features ":"F=S",
5       "POSTag": " substantivo ",
6       "lemmas ":[ " moto " ],
7       "lexeme ":" moto "
8     }
9   ],
10  "TAG ":"NP"
11 }
```

Listing 3.2 – Pseudo-Código de análise sintática em conjunto com gramatical

Com as novas informações geradas é possível utilizar também os lemmas da palavra identificada. Os lemmas encontrados são variações das palavras que estão sendo analisadas. Com essa técnica ao invés de se considerar comparações usando apenas o comparador de igual, é possível utilizar um laço de iteração para comparar as variações da palavra encontrada.

O exemplo abaixo mostra o funcionamento de uma parte do código para realizar a análise utilizando os dados gerados pela análise gramatical e sintática.

```
1 var obj_roubado = [];
2 var carros_palavras = [ 'carro ', 'mercedes ', 'audi ', 'bmw' ]
3
4 para (var i=0; i<tamanho_frase; i++){
5   se(carros_palavras.contem(frase[i] ->palavra)){
6     se(frase[i+1]->POSTag == 'adjetivo '){
7       obj_roubado.adiciona(frase[i]->palavra+' '+frase[i+1]->palavra)
8     }
9   }
10 }
11 ;
```

```
8     } senao {  
9         obj_roubado . adiciona ( frase [ i ] -> palavra ) ;  
10    }  
11 }  
12 }
```

Neste exemplo apresentado é possível perceber que o primeiro "SE utilizado faz uma comparação sintática é conhecida e marcada como carro. Após isso é checado se a palavra seguinte é um adjetivo, utilizando a classe gramatical.

Com a união de técnicas sintáticas e gramaticais a generalização possibilitou grandes avanços na identificação das informações relevantes. Comparações são feitas de maneira rápida e eficiente usando uma programação de alto nível.

Os problemas encontrados neste momento começaram a se apresentar quando era preciso identificar mais informações com mais detalhes. Uma quantidade grande de informações para ser identificada geram grandes trechos de código, o que aumenta rapidamente a complexidade das análises a serem realizadas. Outro problema, em nível de código, identificado nesta parte foi a fácil interferência em casos já descritos enquanto se era criado um novo, já que a dificuldade de identificar como o código se comportaria e até onde ele seria capaz de identificar é grande.

A falta de significado das palavras também pode causar problemas durante esse tipo de análise. Palavras ambíguas podem não ser reconhecidas corretamente nesta seção. Para resolver este problema é preciso a identificação do significado das palavras para auxiliar na identificação dos elementos.

A saída desta análise é um pouco melhor que a saída gerada pela análise puramente sintática. Como resultado desta análise um array chave-valor contém o elemento como chave e o valor corresponde ao que foi encontrado durante a análise.

3.5 Desenvolvimento de Rede Neural para Apoiar a Busca de Elementos Relevantes

Com o aumento de elementos necessário para realizar a identificação a complexidade e dificuldade de criar novos casos para encontrar esses elementos no código aumentou muito rápido, chegando a um ponto em que a manutenção do código começava a se tornar inviável. Um dos resultados gerados, pela dificuldade em codificar, aparecia quando um caso de análise podia estar junto com outro existente, mas isso não tinha sido identificado e acaba gerando dois casos praticamente iguais em locais separados, ou até mesmo quando eu caso novo acaba atrapalhando algum existente. Os problemas com os métodos anteriores não estavam limitados apenas a dificuldade de codificar, mas a velocidade das análises começava a sentir os reflexos

dos grandes trechos de código para cada possível caso.

Com o intuito de resolver os problemas encontrados nas versões anteriores e melhorar as análises o uso de técnicas de aprendizado de máquina foi adotado. Para realizar este trabalho está sendo utilizado uma Rede Neural Artificial (ANN).

Com a utilização de ANN a geração de cada caso de análise pode ser feita de maneira muito mais rápida e automatizada, incluindo novos casos, uma vez que a rede já tenha sido configurada uma vez. Para um novo caso ser adicionado ele passa pelos seguintes passos:

1. Definição do elemento a ser identificado
2. Obter e/ou criar exemplos de frase contendo o novo elemento a ser identificado
3. Separação das partes específicas que fazem a identificação do novo elemento a ser identificado
4. Realizar a análise gramatical e adicionar as informações geradas
5. Converter os dados obtidos para uma posição da entrada da ANN
6. Gerar treinos
7. Treinar ANN
8. Testar e validar com os novos dados

Cada um dos passos acima, apesar de demandarem um certo tempo para a realização, continuam sendo mais rápidos do que montar cada caso separado utilizando apenas análise sintática e gramatical.

Durante a definição de elementos a serem identificados, a enumeração do que falta ser identificado pela análise é feita. Assim que estes elementos são definidos o passo dois é obter textos transcritos ou criar exemplos onde esse elementos apareçam, de preferência de diferentes formas.

Durante a execução do passo 2, deve ser levado em consideração que as análises usaram os exemplos gerados nesse passo para identificar este elemento posteriormente, então, se os casos gerados forem incompletos ou incorretos os resultados da análise pode sofrer com isso.

O passo 3 consiste em pegar os textos obtidos no passo anterior e remover tudo que não é necessário para realizar a identificação do elemento em questão. Sobrando apenas o que é necessário, um array contendo cada um destes casos é gerado. A análise gramatical no passo seguinte pode então adicionar as informações extras necessárias para cada palavra.

Os últimos passos estão diretamente relacionados com a ANN. A conversão dos dados para entradas da ANN deve ocorrer porque a rede utilizada aceita apenas números em sua entrada e os dados disponíveis são textos.

Com a apresentação da imagem 3.4 é possível ver a sequência de passos que o dado obtido deve seguir até o término do processo. Os dados gramaticais, como no passo anterior, continuam ocorrendo logo após a obtenção dos dados. A grande diferença entre este método e o método anterior é a forma como as informações são identificadas. O passo anterior fazia o uso de controles condicionais como "if"s para realizar a identificação dos elementos, este método utiliza uma rede neural que considera os dados sintáticos e gramaticais disponíveis e gera na saída da ANN valores que vão de 0 a 1, representando o quanto o dado usado em sua entrada está próximo de ser um dos elementos relevantes que precisam ser identificados.

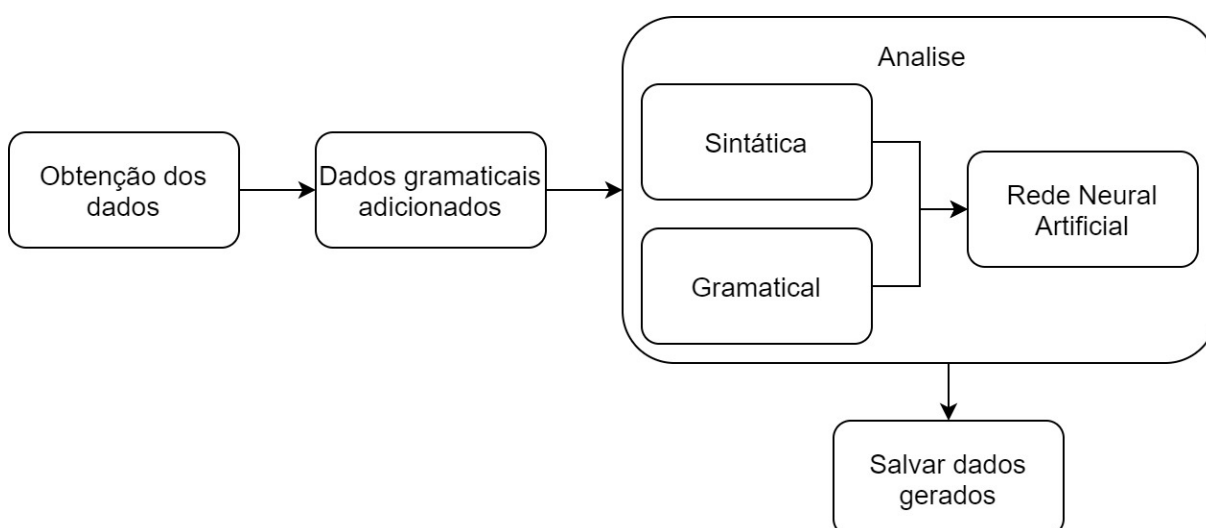


Figura 3.4 – Fluxo de dados durante a identificação de elementos relevantes com ANN

Durante a fase de treino da ANN, todos os exemplos disponíveis e já classificados devem ser utilizados. Durante os testes realizados com os treinos disponíveis cerca de 20 iterações de treino foram o suficiente para deixar a taxa de erros abaixo de 1% nos treinos disponíveis.

Com a facilidade de inserir novos casos a serem identificados, um dos principais problemas gerados até então passa a ser resolvido de forma fácil e eficiente.

A saída desta análise é um array contendo números reais do quanto a entrada está próximo a um elemento relevante da situação, este número vai de 0 a 1. Dessa forma é possível determinar a partir de quando é considerado que o valor usado como entrada realmente corresponde a determinado elemento.

Quando um valor é considerado como correspondente a um elemento, o que foi

usado como entrada para ativação da ANN pode ser salvo em um array chave-valor, com a chave correspondendo ao elemento e o valor o que foi usado como entrada para a ANN.

3.6 Ontologia para Apoiar a Busca de Elementos Relevantes

Com a contínua evolução da pesquisa e o contínuo aumento dos elementos e características de cada elemento que se precisava identificar, a complexidade continuava a aumentar. Com a intenção de melhorar a qualidade dos resultados obtidos da análise o uso de ontologia começou a ser feito. Aliando os dados já usados, como dados sintáticos e gramaticais, com ontologia para adicionar semântica as palavras a rede neural pode utilizar um novo parâmetro para decidir se o que esta sendo analisado realmente é um elemento relevante.

Dessa forma o processo continua o mesmo até chegar a etapa de ativar a rede neural, neste momento, antes de ativar a rede neural, consultas são feitas a ontologia, utilizando SPARQL, e os resultados obtidos são adicionados como uma entrada para a rede neural.

A imagem 3.5 mostra que os dados, após serem obtidos, deve ser classificados sequencialmente gramaticalmente e semanticamente. Com a classificação semântica o significado das palavras já conhecidas pela ontologia é adicionado. A ativação da ANN neste passo leva em consideração para sua ativação, além dos dados usados no passo anterior, a classificação semântica das palavras.

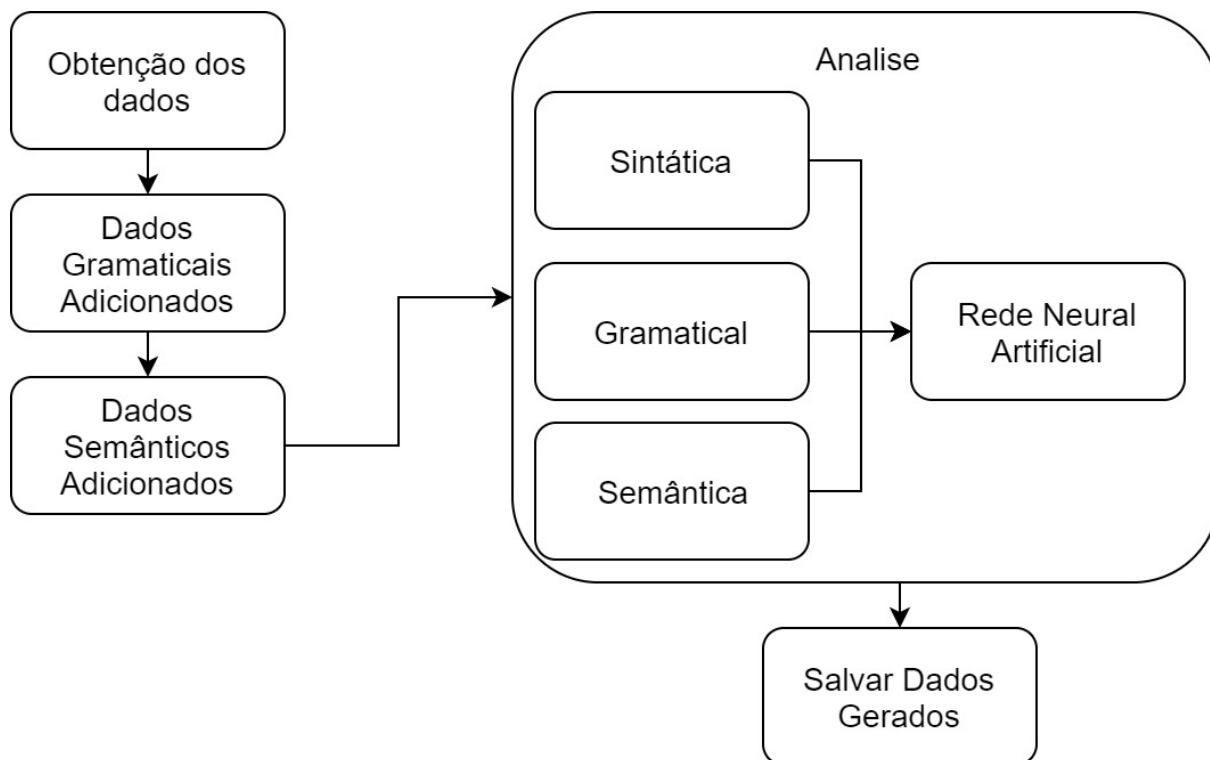


Figura 3.5 – Identificação de elementos relevantes com apoio de ontologia

Com a evolução deste projeto as entradas utilizadas para realizar a classificação e identificação dos elementos relevantes neste passo são:

1. Palavras transcritas
2. Classificação gramatical das palavras
3. Classificação semântica das palavras

Com a utilização dos dados apresentados acima, quando um caso que ainda não é conhecido pela ANN for usado como entrada ela ainda pode reconhecê-lo utilizando os outros dados disponíveis. Quando a entrada contiver uma palavra que não foi utilizada para realizar o treinamento ela ainda pode reconhecê-la utilizando a classificação gramatical ou semântica desta palavra.

O resultado deste método é igual ao obtido sem semântica, a diferença entre estas análises é a consideração da semântica durante a classificação.

3.7 Continuação Pós Identificação de Elementos

Após a classificação do texto, proveniente de rede social ou de fala natural, estar completa, é possível que a continuação do processo para a criação ou manutenção

de SAW aconteça.

Durante o começo desta pesquisa, até a análise gramatical apenas, quando a classificação terminava os dados eram apenas salvos em um banco de dados relacional e a parte seguinte, responsável por calcular a qualidade dessas informações, era notificada.

Com a adição da semântica os dados agora são empacotados dentro de um RDF XML antes de serem enviados para serem salvos. Dessa maneira desde o começo da utilização de ANN os dados são salvos com semântica. Apenas após serem salvos um aviso é enviado para que a próxima parte possa continuar as análises.

Capítulo 4

ESTUDO DE CASO

Durante as seções onde foram analisados os resultados gerados, todos os testes foram realizados dentro de um ambiente simulado e controlado, onde os casos utilizados para realizar as simulações eram conhecidos e foram usados anteriormente para direcionar como as técnicas de análises deveriam funcionar. Esse direcionamento foi feito baseado nos elementos que precisavam ser identificados.

Cada uma das técnicas apresentadas neste trabalho foi testada individualmente e os exemplos utilizados para sua criação, manutenção e utilização podem variar entre os testes. Essa variação acontece devido à evolução natural do trabalho, onde as técnicas testadas mais recentemente tendem a ter mais e melhores casos de teste.

Algumas das frases utilizadas durante a criação das análises e durante os testes estão apresentados abaixo:

"boa noite acaba de acontecer um roubo de carro aqui na domingos setti com a luis vives dois caras de moto apontaram uma arma para o motorista de um mercedes preto e mandaram sair do carro sem levar nada os dois fugiram acelerando em direcao ao metro klabin um deles usava uma camiseta vermelha e o outro tinha um bone preto"

"Socorro um cara foi roubado aqui na minha frente foi na domingos setti perto do restaurante don paladino faz 5 minutos levaram o carro dele e saíram em disparada sentido zona oeste os bandidos tinham revolveres era altos e tinham tatuagens nos bracos"

"ha 2 minutos atras um homem foi ameacado por dois bandidos armados aqui na rua domingos eles deram uma coronhada no senhor e o mandaram sair do carro ele parece estar bem machucado os dois fugiram em direcao a zona oeste levando o carro"

"acabo de ver uma senhora sendo assaltada aqui na praça da sé o meliante levou sua bolsa"

"um cara de camisa azul ameaçou e roubou uma senhora perto da praça da sé ele tinha uma faca"

4.1 Resultados da Análise Sintática

No momento em que os testes utilizando este tipo de análise foram realizados, todos os casos analisados apresentaram uma taxa de 100% de identificação dos elementos necessários para aquele momento. Porém, o objetivo quando esta análise foi testada era constatar a presença de palavras-chave em uma frase, o que é algo bem inferior ao que foi obtido por outras análises.

Abaixo estão listadas 3 denúncias que foram analisadas utilizando apenas a análise sintática. Para a realização desta análise dezenas de casos de teste foram criados utilizando uma linguagem de programação e quando um determinado conjunto era satisfeito o valor usado na análise era salvo.

```
1 {
2   "frase": "boa noite acaba de acontecer um roubo de carro aqui na
   domingos setti com a luis vives dois caras de moto apontaram uma arma
   para o motorista de um mercedes preto e mandaram sair do carro sem levar
   nada os dois fugiram acelerando em direcao ao metro klabin um deles
   usava uma camiseta vermelha e o outro tinha um bone preto",
3   "tipo_crime": ["roubo"],
4   "local": ["domingos setti com a luis vives"],
5   "locais": ["aqui", "metro"],
6   "vestimenta": ["camiseta", "bone"],
7   "carros": ["carro", "mercedes"],
8   "criminoso": ["caras"],
9   "armas": ["arma"]
10 }
11 {
12   "frase": "acabo de ver uma senhora sendo assaltada aqui na praça da se
   o meliante levou sua bolsa",
13   "tipo_crime": ["assaltada"],
14   "local": [],
15   "locais": [],
16   "vestimenta": [],
17   "carros": [],
18   "criminoso": ["meliente"],
19   "armas": []
20 }
21 {
```

```
22 "frase": "um cara de camisa azul ameaçou e roubou uma senhora perto da  
praca da se ele tinha uma faca",  
23 "tipo_crime": [],  
24 "local": [],  
25 "locais": ["praca"],  
26 "vestimenta": ["camisa"],  
27 "carros": [],  
28 "criminoso": [],  
29 "armas": ["faca"]  
30 }
```

Listing 4.1 – Exemplo de resultado gerado pela classificação sintática

A identificação da presença de palavras-chave é uma tarefa simples e funcional, porém, os dados obtidos podem não ser confiáveis sem outras checagens. Casos onde as palavras 'não' e 'tem' aparecem antes da palavra 'arma' pode significar que não existe uma arma em determinada situação, mas a análise exclusivamente sintática pode gerar como resultado apenas que encontrou a palavra 'arma'.

Nota-se que, apesar da alta taxa de sucesso, esta técnica é extremamente primitiva e se aplicada em cenários onde existe um grau maior de complexidade dos dados essa análise tende a ser pouco eficiente. A solução para a insuficiência em cenários complexos pode ser resolvida com a criação de um caso de teste para cada possível variação da informação buscada, que é o mesmo fato que a torna inviável.

4.2 Resultados da Análise Sintática em Conjunto com a Análise Gramatical

Com este tipo de análise o nível dos resultados esperados pode ser aumentado, e a partir desse momento, palavras que não eram conhecidas pelo sistema podem ser corretamente classificadas utilizando a classe gramatical. Por ser uma evolução do método anterior, o que foi avaliado neste momento foi a identificação de elementos e atributos de elementos onde apenas o uso de técnicas de análise sintática e análise gramatical foram usadas. Durante os testes realizados todos os elementos foram identificados com sucesso e seus respectivos atributos também.

No bloco abaixo estão 3 denúncias analisadas utilizando o método sintático auxiliado pelo gramatical. Atributos que ajudam a melhorar a qualidade da informação estão presentes junto com os elementos relevantes identificados.

```
1 {
```

```
2   "frase": "boa noite acaba de acontecer um roubo de carro aqui na
    domingos setti com a luis vives dois caras de moto apontaram uma arma
    para o motorista de um mercedes preto e mandaram sair do carro sem levar
    nada os dois fugiram acelerando em direcao ao metro klabin um deles
    usava uma camiseta vermelha e o outro tinha um bone preto",
3   "tipo_crime": ["roubo"],
4   "local": ["domingos setti com a luis vives"],
5   "locais": ["aqui","metro"],
6   "vestimenta": ["camiseta vermelha","bone preto"],
7   "carros":["carro","mercedes preto"],
8   "criminoso": ["dois caras"],
9   "armas": ["arma"],
10  "tempo": ["acaba acontecer"],
11  "vitima": ["motorista"],
12  "criminoso_fuga": ["fugiram acelerando","metro klabin"],
13  "objeto_roubado": ["bolsa"]
14 }
15 {
16  "frase": "acabo de ver uma senhora sendo assaltada aqui na praca da se
    o meliante levou sua bolsa",
17  "tipo_crime": ["assaltada"],
18  "local": ["praca da se"],
19  "locais": ["praca"],
20  "vestimenta": [],
21  "carros":[],
22  "criminoso": ["meliante"],
23  "armas": [],
24  "tempo": ["acabo de ver"],
25  "vitima": ["senhora"],
26  "criminoso_fuga": ["fugiram acelerando","metro klabin"],
27  "objeto_roubado": ["bolsa"]
28 }
29 {
30  "frase": "um cara de camisa azul ameacou e roubou uma senhora perto da
    praca da se ele tinha uma faca",
31  "tipo_crime": ["roubou"],
32  "local": ["praca da se"],
33  "locais": ["praca"],
34  "vestimenta": ["camisa azul"],
35  "carros":[],
36  "criminoso": [],
37  "armas": ["faca"],
38  "tempo": [],
39  "vitima": ["senhora"],
40  "criminoso_fuga": [],
41  "objeto_roubado": []
42 }
```

Listing 4.2 – Exemplo de resultado gerado pela classificação sintática em conjunto com gramatical

Os resultados obtidos a partir deste tipo de análise são melhores que os obtidos quando comparado com análises exclusivamente sintáticas, já que possibilita um grau maior de generalização e disponibiliza mais dados para realizar as análises.

Apesar de apresentar bons resultados, as vantagens obtidas pela consideração da gramática gera uma complexidade de desenvolvimento muito maior e quando aplicada a cenários onde existem diversos dados complexos para serem identificados pode se tornar inviável.

Para a realização da análise apresentada acima foram utilizadas centenas de casos de teste diferentes. O que gera um grande grau de dificuldade para criação de novos casos e manutenção de casos já existentes. Isso acontece devido a grande quantidade de informações a ser considerada.

4.3 Resultados da Rede Neural Com Dados Sintáticos e Gramaticais

Durante a realização dessa análise foi constatado que o poder de identificação dos elementos relevantes e atributos pode ser aumentado. Logo, o resultado esperado para essa análise era de que classificando automaticamente a entrada, mesmo que faltasse alguma informação, ele fosse capaz de realizar a classificação corretamente.

Na caixa a seguir estão presentes 3 denúncias distintas classificadas utilizando a rede neural juntamente com dados sintáticos e gramaticais. Foram usadas algumas dezenas de partes de denúncias para realizar os treinos antes de poder realmente utilizar a ANN para classificar as informações.

```
1 {  
2   "frase": "boa noite acaba de acontecer um roubo de carro aqui na  
   domingos setti com a luis vives dois caras de moto apontaram uma arma  
   para o motorista de um mercedes preto e mandaram sair do carro sem levar  
   nada os dois fugiram acelerando em direcao ao metro klabin um deles  
   usava uma camiseta vermelha e o outro tinha um bone preto",  
3   "vitima": ["motorista"],  
4   "criminoso": ["dois caras"],  
5   "objeto_roubado": ["mercedes preto"],  
6   "arma": ["uma arma"],  
7   "tempo": ["boa noite","acaba de acontecer"]  
8 }  
9
```

```
10 {
11   "frase": "acabo de ver uma senhora sendo assaltada aqui na praca da se
12   o meliante levou sua bolsa",
13   "vitima": ["senhora"],
14   "criminoso": ["meliante"],
15   "objeto_roubado": ["bolsa"],
16   "arma": [],
17   "tempo": ["acabo de ver"]
18 }
19 {
20   "frase": "um cara de camisa azul ameacou e roubou uma senhora perto da
21   praca da se ele tinha uma faca",
22   "vitima": ["senhora"],
23   "criminoso": ["um cara"],
24   "objeto_roubado": [],
25   "arma": ["faca"],
26   "tempo": []
27 }
```

Listing 4.3 – Exemplo de resultado gerado pela classificação com rede neural

Os resultados encontrados durante os testes apresentaram uma boa identificação de elementos, mas ressaltaram a importância de se treinar a maior quantidade de casos possíveis. Com uma grande quantidade de treinos a classificação pode distinguir a diferença entre frases "não tem arma" e "usou uma arma".

Foi constatado durante os testes que existiram erros durante a classificação de algumas classes, provavelmente ocasionados por falta de casos para treinar a rede neural. A falta de casos para gerar os testes acontece devido a dificuldade em obter denúncias que contenham as informações que precisam ser encontradas. Para melhorar com a falta de exemplos para geração de testes, denúncias fictícias foram geradas. A ocorrência de erros tende a se repetir em momentos que a mesma classe gramatical aparece entre a entrada e alguma classe existente, classificando assim, elementos irrelevantes como relevantes ou deixando de identificar elementos relevantes.

4.4 Resultados da Rede Neural com Dados Sintáticos, Gramaticais e Semânticos

Com o objetivo de resolver os problemas gerados na análise anterior o uso de ontologia foi adotado, juntamente com as outras análises que já eram feitas. Durante os testes realizados e com uma ontologia que contemplava dentro da sua base de conhecimento todas as palavras utilizadas nos testes as análises passaram a ser feita

sem a ocorrência de erros.

Na caixa seguinte estão presentes os resultados das análises de 3 denúncias distintas quando aplicado a técnica que considera dados sintáticos, gramaticais e semânticos utilizando ANN para realizar a classificação. Como a função da semântica é melhorar a qualidade dos dados, a confiabilidade dos dados encontrados abaixo é maior que antes.

```
1 {
2   "frase": "boa noite acaba de acontecer um roubo de carro aqui na
   domingos setti com a luis vives dois caras de moto apontaram uma arma
   para o motorista de um mercedes preto e mandaram sair do carro sem levar
   nada os dois fugiram acelerando em direcao ao metro klabin um deles
   usava uma camiseta vermelha e o outro tinha um bone preto",
3   "vitima": ["motorista"],
4   "criminoso": ["dois caras"],
5   "objeto_roubado": ["mercedes preto"],
6   "arma": ["uma arma"],
7   "tempo": ["boa noite","acaba de acontecer"]
8 }
9
10 {
11   "frase": "acabo de ver uma senhora sendo assaltada aqui na praca da se
   o meliante levou sua bolsa",
12   "vitima": ["senhora"],
13   "criminoso": ["meliante"],
14   "objeto_roubado": ["bolsa"],
15   "arma": [],
16   "tempo": ["acabo de ver"]
17 }
18 {
19   "frase": "um cara de camisa azul ameacou e roubou uma senhora perto da
   praca da se ele tinha uma faca",
20   "vitima": ["senhora"],
21   "criminoso": ["um cara"],
22   "objeto_roubado": [],
23   "arma": ["faca"],
24   "tempo": []
25 }
```

Listing 4.4 – Exemplo de resultado gerado pela classificação com rede neural e dados semânticos

Com a utilização de dados semânticos das palavras, a generalização das classificações pode ser aumentada, já que agora mesmo que um treino com a palavra "faca" nunca tenha sido feito, graças a utilização da ontologia que a classificou como "arma" a classificação da ANN ainda consegue reconhecê-la corretamente.

4.5 Considerações Finais

Durante o desenvolvimento das técnicas ficou claro que a quantidade de informações a serem classificadas e a complexidade das informações influenciam diretamente nos resultados gerados por cada uma das análises. Alguma das análises apresentam melhores resultados com grandes quantidades de dados simples e outras com dados mais complexos.

A tabela 4.1 classifica as técnicas utilizadas para realizar a classificação dos elementos relevantes nas frases transcritas de fala natural, indo de pior (' - -') para melhor ('+ + +').

Tabela 4.1 – Comparação Entre Técnicas para Realização de Entendimento de Fala Natural

	Sintática	+Gramática	+ANN	+Semântica
Criação	+ + +	+	-	-
Manutenção	+ + +	- - -	+ +	+ +
Complexidade Poucos Elementos	+ + +	- -	+ + +	+ + +
Complexidade Muitos Elementos	+	- - -	+ +	+ + +
Qualidade Resultados	- - -	-	+ +	+ + +

Com a análise da tabela apresentada é possível perceber que existem grandes diferenças entre os tipos de análises comparados. A análise puramente sintática apresenta grande facilidade de criação e manutenção, porém, tem dificuldades em identificar completamente os elementos necessários.

A técnica que alia análises sintáticas e gramaticais consegue realizar ter uma melhora no desempenho, quando comparado com a técnica puramente sintática, mas a um alto custo de complexidade de criação e manutenção.

A partir do momento em que algoritmos de machine learning passam a fazer parte das análises a complexidade de manutenção diminuí rapidamente enquanto a de criação tem uma leve piora. Os resultados apresentados por algoritmos que se beneficiam desta técnica tendem a ser melhores quando comparado com os métodos anteriores.

Com a utilização de ontologias para adicionar a semântica de palavras a dificuldade em criar e dar manutenção para esta análise se mantém estável enquanto os resultados tendem a ser melhores, já que podem se beneficiar classificação semântica.

Com os testes realizados, pode ser observado que cada técnica tem pontos

positivos e negativos em determinados casos. A escolha da técnica fica dependente de uma análise feita previamente para obter informações do trabalho a ser realizado, para apenas então, ser decidida.

Como as técnicas apropriadas variam de situação para situação, baseado nos resultados obtidos ao longo da implementação deste trabalho, a imagem 4.1 foi elaborada com objetivo de auxiliar durante a escolha da técnica que deverá ser utilizada em trabalhos futuros. O resultado gerado por este auxílio é uma sugestão, não necessariamente sendo a única alternativa para a situação enfrentada.

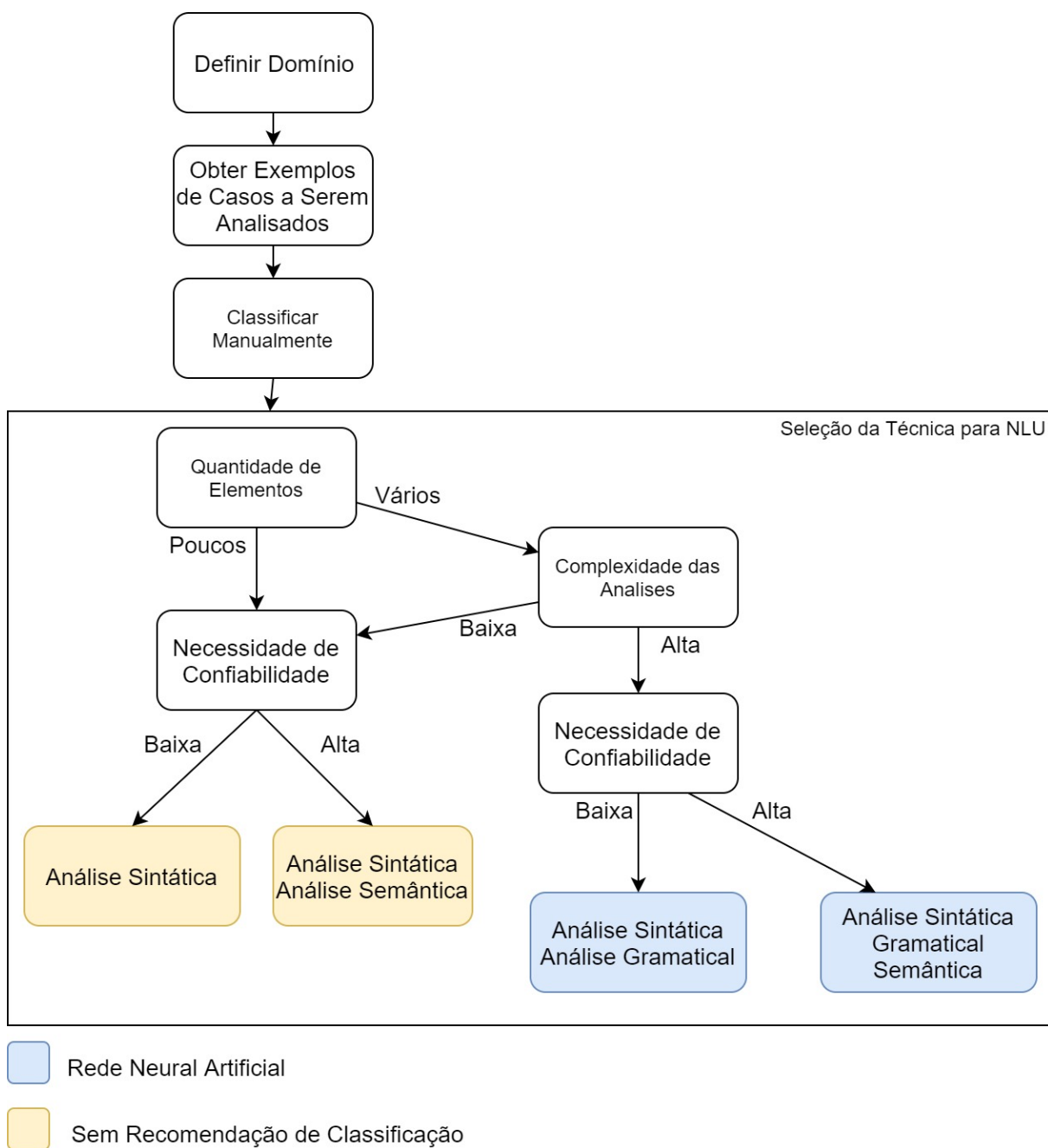


Figura 4.1 – Seleção de técnica para NLU

As caixas coloridas apresentadas na imagem são as técnicas sugeridas para realizar a análise e que devem variar de cenário para cenário. Começando da caixa de definição de cenário e seguindo as setas, pelo caminho que melhor representa o domínio enfrentado, uma técnica para realização do processamento de linguagem natural é encontrada.

A caixa que representada pelo texto "Classificar manualmente" consiste em realizar a análise necessária manualmente, com o objetivo de identificar os elementos que serão necessários identificar e quais atributos eles podem possuir. Com estes dados em mãos é possível dar início a escolha da técnica.

Capítulo 5

CONCLUSÕES

Este trabalho teve como objetivo a procura e identificação de elementos relevantes para o domínio de gerenciamento de emergência criminal com a realização de entendimento de fala natural, para que pudesse ser gerado subsídio, com um alto grau de qualidade, para a geração de SAW. Para isso, foram implementadas técnicas de análise de texto que consideravam desde a forma como uma palavra é escrita até a semântica das palavras. Os dados identificados e classificados semanticamente, podem então, serem utilizados para abastecerem uma ontologia sobre situações de emergência.

Durante o desenvolvimento deste projeto diversas técnicas foram estudadas, implementadas e criadas a fim de alcançar o objetivo de prover melhor subsídio para geração de SAW. Ao se analisar os resultados obtidos ao longo das diferentes implementações foi constatado que existem casos diferentes de quando cada uma das análises apresentadas deve ser utilizada.

Durante a implementação realizada o que mais se aproximou da realização do objetivo deste trabalho foi a união do uso de análise sintática, gramatical e semântica sendo classificadas por uma rede neural artificial.

Apesar do resultado deste trabalho ter sido concluído e apontar que o uso das técnicas descritas usadas em conjunto fornece a melhor solução para este projeto, é importante notar que, a união destas técnicas da maneira apresentada pode não refletir como a melhor escolha em outros projetos com outros elementos a serem identificados.

Como as análises que precisaram ser realizadas foram mudando e aumentando ao longo do projeto foi notado que, a escolha da técnica a ser adotada é direcionada diretamente pelo tipo de informação a ser identificada e a quantidade de informações.

Em projetos onde a necessidade é de se classificar poucos elementos ou que

existam com poucas variações o uso exclusivo da técnica sintática é recomendado, já que é de fácil implementação.

Projetos que tenham uma variedade maior de palavras que podem ser agrupadas pela classe gramatical, como adjetivos e verbos, o uso de análise sintática aliada com a gramatical é recomendado. Vale notar que, a dificuldade de se programar a análise gramatical pode ser grande, e por isso, é recomendado quando a quantidade de elementos não é grande.

O uso de rede neural artificial se mostra interessante quando uma base já classificada existe para a realização dos treinos. Com o uso desse tipo de classificação a dificuldade encontrada na análise gramatical é diminuída.

Com o uso da classificação semântica a qualidade e confiabilidade dos dados aumenta significativamente, assim como, a quantidade de trabalho necessário para preparar a ontologia e leva-la em consideração durante a ativação da rede neural.

Para finalizar, conclui-se que cada domínio pode obter melhores resultados utilizando a técnica apropriada para a realização do entendimento da fala natural. No domínio deste projeto, os melhores resultados foram obtidos com a utilização de análise sintática e gramatical classificadas utilizando ontologia e rede neural artificial. Com a utilização correta das técnicas de análise, foi possível realizar a identificação de elementos e atributos relevantes dentro de textos transcritos da fala de um humano, realizando entediamento de linguagem natural, e gerar a partir disso, informações semânticas sobre uma determinada situação para que se pudesse ser gerado SAW e auxiliar durante a tomada de decisão criminal.

5.1 Publicações

1. Objects Assessment Approach Using Natural Language Processing and Data Quality to Support Emergency Situation Assessment. Sanches, Matheus F.; Junior, Valdir A. P. ; Souza, Jessica O. ; Coneglian, Caio S. ; Jorge, Fábio R. ; Oliveira, Natália P. ; Botega, Leonardo C. . Communications in Computer and Information Science. 617ed.: Springer International Publishing, 2016, v. , p. 238-244.
2. Processamento de Linguagem Natural para a Avaliação de Situações de Emergência com Interface de Realidade Virtual. SANCHES, M. F.; BOTEGA, L. C. ; OLIVEIRA, A. C. M. . Workshop de Realidade Virtual e Aumentada, 2015, Presidente Prudente. Workshop de Realidade Virtual e Aumentada WRVA 2015 - Livro dos Anais, 2015. v. XVI.
3. IDENTIFICAÇÃO DE ATRIBUTOS DE EVENTOS DE ROUBO EM DENÚNCIA

- 190 PARA SISTEMA DE TOMADA DE DECISÃO MILITAR. Sanches, Matheus F.; Botega, Leonardo C. . Conic-Semesp, 2014, São Paulo. Anais do Conic-Semesp / Volume 2, 2014 ? Universidade Cidade de São Paulo ? UNICID, 2014. v. 2.
4. FUSÃO DE NÍVEL 1 UTILIZANDO SEMÂNTICA PARA O DOMÍNIO MILITAR. Sanches, Matheus F.; BOTEAGA, L. C. . VI Congresso de Pesquisa Científica: Inovação, Sustentabilidade, Ética e Cidadania, 2016, Marília. VI Congresso de Pesquisa Científica: Inovação, Sustentabilidade, Ética e Cidadania, 2016. v. 1. p. 56-56.
 5. PROCESSAMENTO DE LINGUAGEM NATURAL APLICADO À AVALIAÇÃO DE SITUAÇÕES DE EMERGÊNCIA. Sanches, Matheus F.; BOTEAGA, L. C. . V Congresso de Pesquisa Científica: Inovação, Sustentabilidade, Ética e Cidadania, 2015. v. 1. Marília. p. 122-122.
 6. IDENTIFICAÇÃO DE ATRIBUTOS DE EVENTOS DE ROUBO EM DENÚNCIA 190 PARA SISTEMA DE TOMADA DE DECISÃO MILITAR. Sanches, Matheus F.; Marília. IV Congresso de Pesquisa Científica: Inovação, Sustentabilidade, Ética e Cidadania, 2014. v. 1. p. 121-121.
 7. Entendimento de Linguagem Natural para a Ampliação da Consciência Situacional de Emergências. Sanches, Matheus F.; BOTEAGA, L. C. . 7º Congresso de Pesquisa Científica - Inovação, Sustentabilidade, Ética e Cidadania, 2017, Marília.
 8. Towards Semantic Fusion Using Information Quality Awareness to Support Emergency Situation Assessment. *Advances*. Junior, Valdir Amancio Pereira ; Sanches, Mathues Ferraroni ; Botega, Leonardo Castro ; Coneglian, Caio Saraiva. *Intelligent Systems and Computing*. 1ed.: Springer International Publishing, 2016, v. 444, p. 145-155.
 9. Using Semantics to Improve Information Fusion and Increase Situational Awareness. Pereira Junior, Valdir A. ; Sanches, Matheus F. ; Botega, Leonardo C. ; Coneglian, Caio S. ; Oliveira, Natália ; Araújo, Regina B. *Advances in Intelligent Systems and Computing*. 491ed.: Springer International Publishing, 2016, v. 617, p. 101-113.
 10. Multi-criteria Fusion of Heterogeneous Information for Improving Situation Awareness on Emergency Management Systems. Pereira, Valdir Amancio ; Sanches, Matheus Ferraroni ; Botega, Leonardo Castro ; Souza, Jessica ; Coneglian, Caio Saraiva ; Fusco, Elvis ; de Campos, Márcio Roberto . *Lecture Notes in Computer Science*. 1ed.: Springer International Publishing, 2015, v. 9173, p. 3-14.

11. Towards semantic fusion using information quality and the assessment of objects and situations to improve emergency situation awareness. PEREIRA, VALDIR A. ; Sanches, Matheus F. ; SARAN, JORDAN F. ; Coneglian, Caio S. ; Botega, Leonardo C. ; DE ARAUJO, REGINA B. . 2016 Eleventh International Conference on Digital Information Management (ICDIM), 2016, Porto. 2016 Eleventh International Conference on Digital Information Management (ICDIM), 2016. v. 1. p. 260-265.

REFERÊNCIAS BIBLIOGRÁFICAS

- AL-HADIDI, M. R.; ALARABEYYAT, A.; ALHANAHNAH, M. Breast cancer detection using k-nearest neighbor machine learning algorithm. In: *Breast Cancer Detection using K-nearest Neighbor Machine Learning Algorithm*. [S.l.: s.n.], 2016.
- BOTEGA, L. C. Modelo de fusão dirigido por humanos e ciente de qualidade de informação. In: . [S.l.: s.n.], 2016.
- CHANG, Y.-W.; TSAI, C.-Y. Apply deep learning neural network to forecast number of tourists. In: *Apply deep learning neural network to forecast number of tourists*. [S.l.: s.n.], 2017.
- CHOWDHURY, G. G. Natural language processing. In: *Natural language processing*. [S.l.: s.n.], 2003.
- COLLOBERT, R.; WESTON, J. A unified architecture for natural language processing: Deep neural networks with multitask learning. In: . [S.l.: s.n.], 2008.
- ENDSLEY, M. R. Toward a theory of situation awareness in dynamic systems. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society 1995 37: 32*. [S.l.: s.n.], 1995.
- ENDSLEY, M. R. Designing for situation awareness in complex system. In: *Proceedings of the Second intentional workshop on symbiosis of humans, artifacts and environment*. [S.l.: s.n.], 2001.
- KINOSHITA, J.; SALVADOR, L. do N.; MENEZES, C. E. D. de. Cogroo – um corretor gramatical para a língua portuguesa,acoplável ao openoffice. In: *CoGrOO – Um Corretor Gramatical para a língua portuguesa,acoplável ao OpenOffice*. [S.l.: s.n.], 2005.
- MATHEUS, C. J.; KOKAR, M. M.; BACLAWSKI, K. A core ontology for situation awareness. In: . [S.l.: s.n.], 2003.
- NOY, N. F.; MCGUINNESS, D. L. Ontology development 101: A guide to creating your first ontology. In: *Ontology Development 101: A Guide to Creating Your First Ontology*. [S.l.: s.n.], 2001.
- PANASETSKY, D.; TOMIN, N. Using of neural network technology and multi-agent systems to preventing large-scale emergencies in electric power systems. In: . [S.l.: s.n.], 2013.

SCHARM, M. et al. Comodi: an ontology to characterise differences in versions of computational models in biology. In: *COMODI: an ontology to characterise differences in versions of computational models in biology*. [S.l.: s.n.], 2016.

SEBASTIANI, F. Machine learning in automated text categorization. In: *Machine Learning in Automated Text Categorization*. [S.l.: s.n.], 2002.

VERMA, S. et al. Natural language processing to the rescue? extracting “situational awareness” tweets during mass emergency. In: . [S.l.: s.n.], 2011.

WINOGRAD, T. Understanding natural language. In: *Understanding Natural Language*. [S.l.: s.n.], 1972.

Apêndice A

TREINO UTILIZADO NA REDE NEURAL ARTIFICIAL

A.1 Arma

```

1 {"sentences":[{"chunks":[{"tokens":[{"features":"F=S","POSTag":"art","lemmas":["um"],"lexeme":"uma"},{"features":"F=S","POSTag":"n","lemmas":["arma"],"lexeme":"arma"}],"TAG":"NP"}],"tokens":[],"syntacticChunks":[],"text":"uma arma"}]}
2 {"sentences":[{"chunks":[{"tokens":[{"features":"F=S","POSTag":"n","lemmas":["arma"],"lexeme":"arma"}],"TAG":"NP"}],"tokens":[],"syntacticChunks":[],"text":"arma"}]}
3 {"sentences":[{"chunks":[{"tokens":[{"features":"M=S","POSTag":"n","lemmas":["revolver"],"lexeme":"revolver"}],"TAG":"NP"}],"tokens":[],"syntacticChunks":[],"text":"revolver"}]}
4 {"sentences":[{"chunks":[{"tokens":[{"features":"F=S","POSTag":"n","lemmas":["pistola"],"lexeme":"pistola"}],"TAG":"NP"}],"tokens":[],"syntacticChunks":[],"text":"pistola"}]}
5 {"sentences":[{"chunks":[{"tokens":[{"features":"F=S","POSTag":"n","lemmas":["faca"],"lexeme":"faca"}],"TAG":"NP"}],"tokens":[],"syntacticChunks":[],"text":"faca"}]}
6 {"sentences":[{"chunks":[{"tokens":[{"features":"F=P","POSTag":"n","lemmas":["pistola"],"lexeme":"pistolas"}],"TAG":"NP"}],"tokens":[],"syntacticChunks":[],"text":"pistolas"}]}
7 {"sentences":[{"chunks":[{"tokens":[{"features":"F=P","POSTag":"n","lemmas":["faca"],"lexeme":"facas"}],"TAG":"NP"}],"tokens":[],"syntacticChunks":[],"text":"facas"}]}
8 {"sentences":[{"chunks":[{"tokens":[{"features":"IMPF=3S=IND","POSTag":"v-fin","lemmas":["ter"],"lexeme":"tinha"}],"TAG":"VP"},{"tokens":[{"features":"F=S","POSTag":"art","lemmas":["um"],"lexeme":"uma"},{"features":"F=S","POSTag":"n","lemmas":["arma"],"lexeme":"arma"}],"TAG":"NP"}],"tokens":[],"syntacticChunks":[],"text":"tinha uma arma"}]}

```

```

9 {"sentences":[{"chunks":[{"tokens":[{"features":"IMPF=3S=IND","POSTag":"v-
  fin","lemmas":["ter"],"lexeme":"tinha"},"TAG":"VP"},{"tokens":[{"
  features":"F=S","POSTag":"n","lemmas":["faca"],"lexeme":"faca"},"TAG":"
  NP"}]},"tokens":[],"syntacticChunks":[],"text":"tinha faca"}]}
10 {"sentences":[{"chunks":[{"tokens":[{"features":"IMPF=3S=IND","POSTag":"v-
  fin","lemmas":["ter"],"lexeme":"tinha"},"TAG":"VP"},{"tokens":[{"
  features":"F=S","POSTag":"n","lemmas":["arma"],"lexeme":"arma"},"TAG":"
  NP"}]},"tokens":[],"syntacticChunks":[],"text":"tinha arma"}]}
11 {"sentences":[{"chunks":[{"tokens":[{"features":"IMPF=3S=IND","POSTag":"v-
  fin","lemmas":["ter"],"lexeme":"tinha"},"TAG":"VP"},{"tokens":[{"
  features":"F=S","POSTag":"art","lemmas":["um"],"lexeme":"uma"},{"
  features":"F=S","POSTag":"n","lemmas":["faca"],"lexeme":"faca"},"TAG":"
  NP"}]},"tokens":[],"syntacticChunks":[],"text":"tinha uma faca"}]}
12 {"sentences":[{"chunks":[{"tokens":[{"features":"IMPF=3S=IND","POSTag":"v-
  fin","lemmas":["ter"],"lexeme":"tinha"},"TAG":"VP"},{"tokens":[{"
  features":"M=S","POSTag":"art","lemmas":["um"],"lexeme":"um"},{"features
  ":"M=S","POSTag":"n","lemmas":["revolver"],"lexeme":"revolver"},"TAG":"
  NP"}]},"tokens":[],"syntacticChunks":[],"text":"tinha um revolver"}]}
13 {"sentences":[{"chunks":[{"tokens":[{"features":"IMPF=3S=IND","POSTag":"v-
  fin","lemmas":["ter"],"lexeme":"tinha"},"TAG":"VP"},{"tokens":[{"
  features":"M=S","POSTag":"n","lemmas":["revolver"],"lexeme":"revolver
 "},"TAG":"NP"}]},"tokens":[],"syntacticChunks":[],"text":"tinha revolver
  "}}]}
14 {"sentences":[{"chunks":[{"tokens":[{"features":"IMPF=3S=IND","POSTag":"v-
  fin","lemmas":["ter"],"lexeme":"tinha"},"TAG":"VP"},{"tokens":[{"
  features":"F=S","POSTag":"n","lemmas":[],"lexeme":"pistola"},"TAG":"NP
  "}},"tokens":[],"syntacticChunks":[],"text":"tinha pistola"}]}
15 {"sentences":[{"chunks":[{"tokens":[{"features":"IMPF=3S=IND","POSTag":"v-
  fin","lemmas":["ter"],"lexeme":"tinha"},"TAG":"VP"},{"tokens":[{"
  features":"F=S","POSTag":"art","lemmas":["um"],"lexeme":"uma"},{"
  features":"F=S","POSTag":"n","lemmas":["pistola"],"lexeme":"pistola"},"
  TAG":"NP"}]},"tokens":[],"syntacticChunks":[],"text":"tinha uma pistola
  "}}]}
16 {"sentences":[{"chunks":[{"tokens":[{"features":"M=S","POSTag":"v-pcp","
  lemmas":["armar"],"lexeme":"armado"},"TAG":"VP"}]},"tokens":[],"
  syntacticChunks":[],"text":"armado"}]}
17 {"sentences":[{"chunks":[{"tokens":[{"features":"M=S","POSTag":"n","lemmas
  ":" tiro"],"lexeme":" tiro"},"TAG":"NP"}]},"tokens":[],"syntacticChunks
  ":"[],"text":" tiro"}]}
18 {"sentences":[{"chunks":[{"tokens":[{"features":"M=S","POSTag":"n","lemmas
  ":" tiroteio"],"lexeme":" tiroteio"},"TAG":"NP"}]},"tokens":[],"
  syntacticChunks ":"[],"text":" tiroteio"}]}
19 {"sentences":[{"chunks":[{"tokens":[{"features":"M=P","POSTag":"v-pcp","
  lemmas":["armar"],"lexeme":"armados"},"TAG":"VP"}]},"tokens":[],"
  syntacticChunks ":"[],"text":"armados"}]}

```

A.2 Criminoso

```

1 {"sentences":[{"chunks":[{"tokens":[{"features":"F=P","POSTag":"n","lemmas":"cara","lexeme":"caras"}],"TAG":"NP"},{"tokens":[{"features":"MQP=3P=IND","POSTag":"v-fin","lemmas":["apontar"],"lexeme":"apontaram"}],"TAG":"VP"},{"tokens":[{"features":"F=S","POSTag":"art","lemmas":["um"],"lexeme":"uma"},{"features":"F=S","POSTag":"n","lemmas":["arma"],"lexeme":"arma"}],"TAG":"NP"}],"tokens":[],"syntacticChunks":[],"text":"caras apontaram uma arma"}]}
2 {"sentences":[{"chunks":[{"tokens":[{"features":"F=P","POSTag":"n","lemmas":"cara","lexeme":"caras"}],"TAG":"NP"},{"tokens":[{"features":"MQP=3P=IND","POSTag":"v-fin","lemmas":["atirar"],"lexeme":"atiraram"}],"TAG":"VP"}],"tokens":[],"syntacticChunks":[],"text":"caras atiraram"}]}
3 {"sentences":[{"chunks":[{"tokens":[{"features":"F=P","POSTag":"n","lemmas":"cara","lexeme":"caras"}],"TAG":"NP"},{"tokens":[{"features":"MQP=3P=IND","POSTag":"v-fin","lemmas":["bater"],"lexeme":"bateram"}],"TAG":"VP"}],"tokens":[],"syntacticChunks":[],"text":"caras bateram"}]}
4 {"sentences":[{"chunks":[{"tokens":[{"features":"F=P","POSTag":"n","lemmas":"cara","lexeme":"caras"}],"TAG":"NP"},{"tokens":[{"features":"MQP=3P=IND","POSTag":"v-fin","lemmas":["roubar"],"lexeme":"roubaram"}],"TAG":"VP"}],"tokens":[],"syntacticChunks":[],"text":"caras roubaram"}]}
5 {"sentences":[{"chunks":[{"tokens":[{"features":"F=P","POSTag":"n","lemmas":"cara","lexeme":"caras"},{"features":"M=P","POSTag":"v-pcp","lemmas":["armar"],"lexeme":"armados"}],"TAG":"NP"}],"tokens":[],"syntacticChunks":[],"text":"caras armados"}]}
6 {"sentences":[{"chunks":[{"tokens":[{"features":"F=P","POSTag":"n","lemmas":"cara","lexeme":"caras"}],"TAG":"NP"},{"tokens":[{"features":"MQP=3P=IND","POSTag":"v-fin","lemmas":["acabar"],"lexeme":"acabaram"},{"features":"-","POSTag":"prp","lemmas":["de"],"lexeme":"de"},{"features":"-","POSTag":"v-inf","lemmas":["roubar"],"lexeme":"roubar"}],"TAG":"VP"}],"tokens":[],"syntacticChunks":[],"text":"caras acabaram de roubar"}]}
7 {"sentences":[{"chunks":[{"tokens":[{"features":"F=P","POSTag":"n","lemmas":"cara","lexeme":"caras"}],"TAG":"NP"},{"tokens":[{"features":"MQP=3P=IND","POSTag":"v-fin","lemmas":["acabar"],"lexeme":"acabaram"},{"features":"-","POSTag":"prp","lemmas":["de"],"lexeme":"de"},{"features":"-","POSTag":"v-inf","lemmas":["assaltar"],"lexeme":"assaltar"}],"TAG":"VP"}],"tokens":[],"syntacticChunks":[],"text":"caras acabaram de assaltar"}]}
8 {"sentences":[{"chunks":[{"tokens":[{"features":"F=S","POSTag":"n","lemmas":"cara","lexeme":"cara"}],"TAG":"NP"},{"tokens":[{"features":"PS=3S=IND","POSTag":"v-fin","lemmas":["apontar"],"lexeme":"apontou"}],"TAG":"VP"},{"tokens":[{"features":"F=S","POSTag":"art","lemmas":["um"],"lexeme":"uma"},{"features":"F=S","POSTag":"n","lemmas":["arma"],"lexeme":"arma"}],"TAG":"NP"}],"tokens":[],"syntacticChunks":[],"text":"cara apontou uma arma"}]}
9 {"sentences":[{"chunks":[{"tokens":[{"features":"F=S","POSTag":"n","lemmas":"cara","lexeme":"cara"},{"features":"M=S","POSTag":"v-pcp","lemmas":["armar"],"lexeme":"armado"}],"TAG":"NP"}],"tokens":[],"syntacticChunks":

```

```

    ":[], "text": "cara armado"]}]
10 {"sentences": [{"chunks": [{"tokens": [{"features": "F=S", "POSTag": "n", "lemmas":
    ["cara"], "lexeme": "cara"}], "TAG": "NP"}, {"tokens": [{"features": "PS=3S=
    IND", "POSTag": "v-fin", "lemmas": ["roubar"], "lexeme": "roubou"}], "TAG": "VP
    "}], "tokens": [], "syntacticChunks": [], "text": "cara roubou"}]}
11 {"sentences": [{"chunks": [{"tokens": [{"features": "F=S", "POSTag": "n", "lemmas":
    ["cara"], "lexeme": "cara"}], "TAG": "NP"}, {"tokens": [{"features": "-", "
    POSTag": "v-ger", "lemmas": ["atirar"], "lexeme": "atirando"}], "TAG": "VP"}], "
    tokens": [], "syntacticChunks": [], "text": "cara atirando"}]}
12 {"sentences": [{"chunks": [{"tokens": [{"features": "F=S", "POSTag": "n", "lemmas":
    ["cara"], "lexeme": "cara"}], "TAG": "NP"}, {"tokens": [{"features": "-", "
    POSTag": "v-ger", "lemmas": ["roubar"], "lexeme": "roubando"}], "TAG": "VP"}], "
    tokens": [], "syntacticChunks": [], "text": "cara roubando"}]}
13 {"sentences": [{"chunks": [{"tokens": [{"features": "F=S", "POSTag": "n", "lemmas":
    ["cara"], "lexeme": "cara"}], "TAG": "NP"}, {"tokens": [{"features": "PR=3S=
    IND", "POSTag": "v-fin", "lemmas": ["acabar"], "lexeme": "acaba"}, {"features":
    "-", "POSTag": "prp", "lemmas": ["de"], "lexeme": "de"}, {"features": "-", "
    POSTag": "v-inf", "lemmas": ["roubar"], "lexeme": "roubar"}], "TAG": "VP"}], "
    tokens": [], "syntacticChunks": [], "text": "cara acaba de roubar"}]}
14 {"sentences": [{"chunks": [{"tokens": [{"features": "F=S", "POSTag": "n", "lemmas":
    ["cara"], "lexeme": "cara"}], "TAG": "NP"}, {"tokens": [{"features": "PR=3S=
    IND", "POSTag": "v-fin", "lemmas": ["roubar"], "lexeme": "rouba"}], "TAG": "VP
    "}], "tokens": [], "syntacticChunks": [], "text": "cara rouba"}]}
15 {"sentences": [{"chunks": [{"tokens": [{"features": "F=S", "POSTag": "n", "lemmas":
    ["cara"], "lexeme": "cara"}], "TAG": "NP"}, {"tokens": [{"features": "PR=3S=
    IND", "POSTag": "v-fin", "lemmas": ["assaltar"], "lexeme": "assalta"}], "TAG": "
    VP"}], "tokens": [], "syntacticChunks": [], "text": "cara assalta"}]}
16 {"sentences": [{"chunks": [{"tokens": [{"features": "F=S", "POSTag": "n", "lemmas":
    ["cara"], "lexeme": "cara"}], "TAG": "NP"}, {"tokens": [{"features": "PR=3S=
    IND", "POSTag": "v-fin", "lemmas": ["acabar"], "lexeme": "acaba"}, {"features":
    "-", "POSTag": "prp", "lemmas": ["de"], "lexeme": "de"}, {"features": "-", "
    POSTag": "v-inf", "lemmas": ["assaltar"], "lexeme": "assaltar"}], "TAG": "VP
    "}], "tokens": [], "syntacticChunks": [], "text": "cara acaba de assaltar"}]}
17 {"sentences": [{"chunks": [{"tokens": [{"features": "M=S", "POSTag": "n", "lemmas":
    ["bandido"], "lexeme": "bandido"}], "TAG": "NP"}], "tokens": [], "
    syntacticChunks": [], "text": "bandido"}]}
18 {"sentences": [{"chunks": [{"tokens": [{"features": "M=S", "POSTag": "n", "lemmas":
    ["bandido"], "lexeme": "bandido"}, {"features": "M=S", "POSTag": "v-pcp", "
    lemmas": ["armar"], "lexeme": "armado"}], "TAG": "NP"}], "tokens": [], "
    syntacticChunks": [], "text": "bandido armado"}]}
19 {"sentences": [{"chunks": [{"tokens": [{"features": "M=P", "POSTag": "n", "lemmas":
    ["bandido"], "lexeme": "bandidos"}, {"features": "M=P", "POSTag": "v-pcp", "
    lemmas": ["armar"], "lexeme": "armados"}], "TAG": "NP"}], "tokens": [], "
    syntacticChunks": [], "text": "bandidos armados"}]}
20 {"sentences": [{"chunks": [{"tokens": [{"features": "M=S", "POSTag": "n", "lemmas":
    ["ladrao"], "lexeme": "ladrao"}], "TAG": "NP"}], "tokens": [], "
    syntacticChunks": [], "text": "ladrao"}]}

```

```

21 {"sentences":[{"chunks":[{"tokens":[{"features":"M=S","POSTag":"n","lemmas":["ladrao"],"lexeme":"ladrao"}, {"features":"M=S","POSTag":"v-pcp","lemmas":["armar"],"lexeme":"armado"}],"TAG":"NP"}],"tokens":[],"syntacticChunks":[],"text":"ladrao armado"}}]
22 {"sentences":[{"chunks":[{"tokens":[{"features":"M=S","POSTag":"adj","lemmas":["criminoso"],"lexeme":"criminoso"}],"TAG":"NP"}],"tokens":[],"syntacticChunks":[],"text":"criminoso"}}]
23 {"sentences":[{"chunks":[{"tokens":[{"features":"M=S","POSTag":"n","lemmas":["homem"],"lexeme":"homem"}],"TAG":"NP"}, {"tokens":[{"features":"-","POSTag":"prp","lemmas":["com"],"lexeme":"com"}],"TAG":"PP"}, {"tokens":[{"features":"F=S","POSTag":"art","lemmas":["um"],"lexeme":"uma"}, {"features":"F=S","POSTag":"n","lemmas":["arma"],"lexeme":"arma"}],"TAG":"NP"}],"tokens":[],"syntacticChunks":[],"text":"homem com uma arma"}}]
24 {"sentences":[{"chunks":[{"tokens":[{"features":"M=S","POSTag":"n","lemmas":["homem"],"lexeme":"homem"}],"TAG":"NP"}, {"tokens":[{"features":"PR=3S=IND","POSTag":"v-fin","lemmas":["roubar"],"lexeme":"rouba"}],"TAG":"VP"}, {"tokens":[{"features":"M=S","POSTag":"n","lemmas":["carro"],"lexeme":"carro"}],"TAG":"NP"}],"tokens":[],"syntacticChunks":[],"text":"homem rouba carro"}}]

```

A.3 Objeto

```

1 {"sentences":[{"chunks":[{"tokens":[{"features":"PS=3P=IND","POSTag":"v-fin","lemmas":["levar"],"lexeme":"levaram"}],"TAG":"VP"}, {"tokens":[{"features":"M=S","POSTag":"art","lemmas":["o"],"lexeme":"o"}, {"features":"M=S","POSTag":"n","lemmas":["carro"],"lexeme":"carro"}],"TAG":"NP"}],"tokens":[],"syntacticChunks":[],"text":"levaram o carro"}}]
2 {"sentences":[{"chunks":[{"tokens":[{"features":"M=S","POSTag":"n","lemmas":["roubo"],"lexeme":"roubo"}],"TAG":"NP"}, {"tokens":[{"features":"-","POSTag":"prp","lemmas":["de"],"lexeme":"de"}],"TAG":"PP"}, {"tokens":[{"features":"M=S","POSTag":"n","lemmas":["carro"],"lexeme":"carro"}],"TAG":"NP"}],"tokens":[],"syntacticChunks":[],"text":"roubo de carro"}}]
3 {"sentences":[{"chunks":[{"tokens":[{"features":"PS=3P=IND","POSTag":"v-fin","lemmas":["roubar"],"lexeme":"roubaram"}],"TAG":"VP"}, {"tokens":[{"features":"M=S","POSTag":"art","lemmas":["o"],"lexeme":"o"}, {"features":"M=S","POSTag":"n","lemmas":["carro"],"lexeme":"carro"}],"TAG":"NP"}],"tokens":[],"syntacticChunks":[],"text":"roubaram o carro"}}]
4 {"sentences":[{"chunks":[{"tokens":[{"features":"M=S","POSTag":"n","lemmas":["carro"],"lexeme":"carro"}],"TAG":"NP"}, {"tokens":[{"features":"PS=3S=IND","POSTag":"v-fin","lemmas":["ir","ser"],"lexeme":"foi"}, {"features":"M=S","POSTag":"v-pcp","lemmas":["roubar"],"lexeme":"roubado"}],"TAG":"VP"}],"tokens":[],"syntacticChunks":[],"text":"carro foi roubado"}}]
5 {"sentences":[{"chunks":[{"tokens":[{"features":"M=S","POSTag":"n","lemmas":["carro"],"lexeme":"carro"}, {"features":"M=S","POSTag":"v-pcp","lemmas":["roubar"],"lexeme":"roubado"}],"TAG":"NP"}],"tokens":[],"syntacticChunks":[],"text":"carro roubado"}}]

```

```

6 {"sentences":[{"chunks":[{"tokens":[{"features":"MQP=3P=IND","POSTag":"v-
  fin","lemmas":["pegar"],"lexeme":"pegaram"}],"TAG":"VP"},{"tokens":[{"
  features":"M=S","POSTag":"pron-det","lemmas":["todo"],"lexeme":"todo
  "},"{"features":"M=S","POSTag":"n","lemmas":["dinheiro"],"lexeme":"
  dinheiro"}],"TAG":"NP"}],"tokens":[],"syntacticChunks":[],"text":"pegaram
  todo dinheiro"}}}
7 {"sentences":[{"chunks":[{"tokens":[{"features":"PR=3S=IND","POSTag":"v-fin
  ","lemmas":["roubar"],"lexeme":"rouba"}],"TAG":"VP"},{"tokens":[{"
  features":"M=S","POSTag":"n","lemmas":["carro"],"lexeme":"carro"}],"TAG
  ":"NP"}],"tokens":[],"syntacticChunks":[],"text":"rouba carro"}}}
8 {"sentences":[{"chunks":[{"tokens":[{"features":"PS=3S=IND","POSTag":"v-fin
  ","lemmas":["roubar"],"lexeme":"roubou"}],"TAG":"VP"},{"tokens":[{"
  features":"M=S","POSTag":"n","lemmas":["carro"],"lexeme":"carro"}],"TAG
  ":"NP"}],"tokens":[],"syntacticChunks":[],"text":"roubou carro"}}}
9 {"sentences":[{"chunks":[{"tokens":[{"features":"PS=3S=IND","POSTag":"v-fin
  ","lemmas":["roubar"],"lexeme":"roubou"}],"TAG":"VP"},{"tokens":[{"
  features":"M=S","POSTag":"art","lemmas":["um"],"lexeme":"um"},"{"features
  ":"M=S","POSTag":"n","lemmas":["carro"],"lexeme":"carro"}],"TAG":"NP
  "}}],"tokens":[],"syntacticChunks":[],"text":"roubou um carro"}}}
10 {"sentences":[{"chunks":[{"tokens":[{"features":"PS=3S=IND","POSTag":"v-fin
  ","lemmas":["roubar"],"lexeme":"roubou"}],"TAG":"VP"},{"tokens":[{"
  features":"M=S","POSTag":"art","lemmas":["o"],"lexeme":"o"},"{"features
  ":"M=S","POSTag":"n","lemmas":["carro"],"lexeme":"carro"}],"TAG":"NP
  "}}],"tokens":[],"syntacticChunks":[],"text":"roubou o carro"}}}
11 {"sentences":[{"chunks":[{"tokens":[{"features":"M=S","POSTag":"n","lemmas
  ":"carro"],"lexeme":"carro"}],"TAG":"NP"},{"tokens":[{"features":"PR=3S
  =IND","POSTag":"v-fin","lemmas":["ser"],"lexeme":"e"},"{"features":"M=S
  ","POSTag":"v-pcp","lemmas":["roubar"],"lexeme":"roubado"}],"TAG":"VP
  "}}],"tokens":[],"syntacticChunks":[],"text":"carro e roubado"}}}

```

A.4 Tempo

```

1 {"sentences":[{"chunks":[{"tokens":[{"features":"PR=3S=IND","POSTag":"v-fin
  ","lemmas":["fazer"],"lexeme":"faz"}],"TAG":"VP"},{"tokens":[{"features
  ":"M=S","POSTag":"art","lemmas":["um"],"lexeme":"um"},"{"features":"M=S
  ","POSTag":"n","lemmas":["minuto"],"lexeme":"minuto"}],"TAG":"NP"}],"
  tokens":[],"syntacticChunks":[],"text":"faz um minuto"}}}
2 {"sentences":[{"chunks":[{"tokens":[{"features":"PR=3S=IND","POSTag":"v-fin
  ","lemmas":["fazer"],"lexeme":"faz"}],"TAG":"VP"},{"tokens":[{"features
  ":"M=P","POSTag":"num","lemmas":["dois"],"lexeme":"dois"},"{"features":"M
  =P","POSTag":"n","lemmas":["minuto"],"lexeme":"minutos"}],"TAG":"NP"}],"
  tokens":[],"syntacticChunks":[],"text":"faz dois minutos"}}}
3 {"sentences":[{"chunks":[{"tokens":[{"features":"PR=3S=IND","POSTag":"v-fin
  ","lemmas":["fazer"],"lexeme":"faz"}],"TAG":"VP"},{"tokens":[{"features
  ":"M=P","POSTag":"num","lemmas":["tres"],"lexeme":"tres"},"{"features":"M

```



```

    =P", "POSTag": "n", "lemmas": ["minuto"], "lexeme": "minutos"}], "TAG": "NP"}], "
    tokens": [], "syntacticChunks": [], "text": "faz tres minutos"}}}
4 {"sentences": [{"chunks": [{"tokens": [{"features": "PR=3S=IND", "POSTag": "v-fin
    ", "lemmas": ["fazer"], "lexeme": "faz"}], "TAG": "VP"}, {"tokens": [{"features
    ": "M=P", "POSTag": "num", "lemmas": ["quatro"], "lexeme": "quatro"}, {"features
    ": "M=P", "POSTag": "n", "lemmas": ["minuto"], "lexeme": "minutos"}], "TAG": "NP
    "}], "tokens": [], "syntacticChunks": [], "text": "faz quatro minutos"}}}
5 {"sentences": [{"chunks": [{"tokens": [{"features": "PR=3S=IND", "POSTag": "v-fin
    ", "lemmas": ["fazer"], "lexeme": "faz"}], "TAG": "VP"}, {"tokens": [{"features
    ": "M=P", "POSTag": "num", "lemmas": ["cinco"], "lexeme": "cinco"}, {"features
    ": "M=P", "POSTag": "n", "lemmas": ["minuto"], "lexeme": "minutos"}], "TAG": "NP
    "}], "tokens": [], "syntacticChunks": [], "text": "faz cinco minutos"}}}
6 {"sentences": [{"chunks": [{"tokens": [{"features": "PR=3S=IND", "POSTag": "v-fin
    ", "lemmas": ["haver"], "lexeme": "ha"}], "TAG": "VP"}, {"tokens": [{"features
    ": "M=S", "POSTag": "num", "lemmas": ["um"], "lexeme": "um"}, {"features": "M=S
    ", "POSTag": "n", "lemmas": ["minuto"], "lexeme": "minuto"}], "TAG": "NP"}], "
    tokens": [], "syntacticChunks": [], "text": "ha um minuto"}}}
7 {"sentences": [{"chunks": [{"tokens": [{"features": "PR=3S=IND", "POSTag": "v-fin
    ", "lemmas": ["haver"], "lexeme": "ha"}], "TAG": "VP"}, {"tokens": [{"features
    ": "M=P", "POSTag": "num", "lemmas": ["dois"], "lexeme": "dois"}, {"features": "M
    =P", "POSTag": "n", "lemmas": ["minuto"], "lexeme": "minutos"}], "TAG": "NP"}], "
    tokens": [], "syntacticChunks": [], "text": "ha dois minutos"}}}
8 {"sentences": [{"chunks": [{"tokens": [{"features": "PR=3S=IND", "POSTag": "v-fin
    ", "lemmas": ["haver"], "lexeme": "ha"}], "TAG": "VP"}, {"tokens": [{"features
    ": "M=P", "POSTag": "num", "lemmas": ["tres"], "lexeme": "tres"}, {"features": "M
    =P", "POSTag": "n", "lemmas": ["minuto"], "lexeme": "minutos"}], "TAG": "NP"}], "
    tokens": [], "syntacticChunks": [], "text": "ha tres minutos"}}}
9 {"sentences": [{"chunks": [{"tokens": [{"features": "PR=3S=IND", "POSTag": "v-fin
    ", "lemmas": ["haver"], "lexeme": "ha"}], "TAG": "VP"}, {"tokens": [{"features
    ": "M=P", "POSTag": "num", "lemmas": ["quatro"], "lexeme": "quatro"}, {"features
    ": "M=P", "POSTag": "n", "lemmas": ["minuto"], "lexeme": "minutos"}], "TAG": "NP
    "}], "tokens": [], "syntacticChunks": [], "text": "ha quatro minutos"}}}
10 {"sentences": [{"chunks": [{"tokens": [{"features": "PR=3S=IND", "POSTag": "v-fin
    ", "lemmas": ["haver"], "lexeme": "ha"}], "TAG": "VP"}, {"tokens": [{"features
    ": "M=P", "POSTag": "num", "lemmas": ["cinco"], "lexeme": "cinco"}, {"features
    ": "M=P", "POSTag": "n", "lemmas": ["minuto"], "lexeme": "minutos"}], "TAG": "NP
    "}], "tokens": [], "syntacticChunks": [], "text": "ha cinco minutos"}}}
11 {"sentences": [{"chunks": [{"tokens": [{"features": "PR=3S=IND", "POSTag": "v-fin
    ", "lemmas": ["acabar"], "lexeme": "acaba"}], "TAG": "VP"}, {"tokens": [{"featu
    res": "-", "POSTag": "prp", "lemmas": ["de"], "lexeme": "de"}, {"features
    ": "-", "POSTag": "v-inf", "lemmas": ["acontecer"], "lexeme": "acontecer"}], "
    TAG": "PP"}], "tokens": [], "syntacticChunks": [], "text": "acaba de acontecer
    "}}
12 {"sentences": [{"chunks": [{"tokens": [{"features": "-", "POSTag": "adv", "lemmas
    ": ["agora"], "lexeme": "agora"}], "TAG": "ADVP"}, {"tokens": [{"features
    ": "-", "POSTag": "prp", "lemmas": ["a"], "lexeme": "a"}, {"features": "-", "
    POSTag": "adv", "lemmas": ["pouco"], "lexeme": "pouco"}], "TAG": "PP"}], "tokens
  
```

```

    ":[], "syntacticChunks": [], "text": "agora a pouco"}}}
13 {"sentences": [{"chunks": [{"tokens": [{"features": "M=S", "POSTag": "pron-det", "lemmas": ["pouco"], "lexeme": "pouco"}, {"features": "M=S", "POSTag": "n", "lemmas": ["tempo"], "lexeme": "tempo"}], "TAG": "NP"}], "tokens": [], "syntacticChunks": [], "text": "pouco tempo"}}}
14 {"sentences": [{"chunks": [{"tokens": [{"features": "PR=1S=IND", "POSTag": "v-fin", "lemmas": ["acabar"], "lexeme": "acabo"}], "TAG": "VP"}, {"tokens": [{"features": "-", "POSTag": "prp", "lemmas": ["de"], "lexeme": "de"}, {"features": "-", "POSTag": "v-inf", "lemmas": ["ver"], "lexeme": "ver"}], "TAG": "PP"}], "tokens": [], "syntacticChunks": [], "text": "acabo de ver"}}}
15 {"sentences": [{"chunks": [{"tokens": [{"features": "PR=3S=IND", "POSTag": "v-fin", "lemmas": ["acabar"], "lexeme": "acaba"}], "TAG": "VP"}, {"tokens": [{"features": "-", "POSTag": "prp", "lemmas": ["de"], "lexeme": "de"}, {"features": "-", "POSTag": "v-inf", "lemmas": ["acontecer"], "lexeme": "acontecer"}], "TAG": "PP"}], "tokens": [], "syntacticChunks": [], "text": "acaba de acontecer"}}}

```

A.5 Vitima

```

1 {"sentences": [{"chunks": [{"tokens": [{"features": "F=S", "POSTag": "n", "lemmas": ["cara"], "lexeme": "cara"}], "TAG": "NP"}, {"tokens": [{"features": "PS=3S=IND", "POSTag": "v-fin", "lemmas": ["ir", "ser"], "lexeme": "foi"}, {"features": "M=S", "POSTag": "v-pcp", "lemmas": ["roubar"], "lexeme": "roubado"}], "TAG": "VP"}], "tokens": [], "syntacticChunks": [], "text": "cara foi roubado"}}}
2 {"sentences": [{"chunks": [{"tokens": [{"features": "M=S", "POSTag": "n", "lemmas": ["senhor"], "lexeme": "senhor"}], "TAG": "NP"}, {"tokens": [{"features": "PS=3S=IND", "POSTag": "v-fin", "lemmas": ["ir", "ser"], "lexeme": "foi"}, {"features": "M=S", "POSTag": "v-pcp", "lemmas": ["roubar"], "lexeme": "roubado"}], "TAG": "VP"}], "tokens": [], "syntacticChunks": [], "text": "senhor foi roubado"}}}
3 {"sentences": [{"chunks": [{"tokens": [{"features": "F=S", "POSTag": "n", "lemmas": ["senhor"], "lexeme": "senhora"}], "TAG": "NP"}, {"tokens": [{"features": "PS=3S=IND", "POSTag": "v-fin", "lemmas": ["ir", "ser"], "lexeme": "foi"}, {"features": "F=S", "POSTag": "v-pcp", "lemmas": ["roubar"], "lexeme": "roubada"}], "TAG": "VP"}], "tokens": [], "syntacticChunks": [], "text": "senhora foi roubada"}}}
4 {"sentences": [{"chunks": [{"tokens": [{"features": "M=S", "POSTag": "art", "lemmas": ["um"], "lexeme": "um"}, {"features": "M=S", "POSTag": "n", "lemmas": ["homem"], "lexeme": "homem"}], "TAG": "NP"}, {"tokens": [{"features": "PS=3S=IND", "POSTag": "v-fin", "lemmas": ["ir", "ser"], "lexeme": "foi"}, {"features": "M=S", "POSTag": "v-pcp", "lemmas": ["ameacar"], "lexeme": "ameacado"}], "TAG": "VP"}], "tokens": [], "syntacticChunks": [], "text": "um homem foi ameacado"}}}
5 {"sentences": [{"chunks": [{"tokens": [{"features": "F=S", "POSTag": "art", "lemmas": ["um"], "lexeme": "uma"}, {"features": "F=S", "POSTag": "n", "lemmas": ["senhor"], "lexeme": "senhora"}], "TAG": "NP"}, {"tokens": [{"features": "PS=3S=IND", "POSTag": "v-fin", "lemmas": ["ir", "ser"], "lexeme": "foi"}, {"

```

```

features ":"F=S", "POSTag":"v-pcp", "lemmas":["ameacar"], "lexeme":"ameacada
"}} , "TAG":"VP"}}, {"tokens":[] , "syntacticChunks":[] , "text":"uma senhora foi
ameacada "}}
6 {"sentences":[{"chunks":[{"tokens":[{"features":"F=S", "POSTag":"n", "lemmas
":["cara"], "lexeme":"cara"}], "TAG":"NP"}], {"tokens":[{"features":"PS=3S=
IND", "POSTag":"v-fin", "lemmas":["acabar"], "lexeme":"acabou"}, {"features
":"-", "POSTag":"prp", "lemmas":["de"], "lexeme":"de"}, {"features":"-", "
POSTag":"v-inf", "lemmas":["ser"], "lexeme":"ser"}, {"features":"M=S", "
POSTag":"v-pcp", "lemmas":["roubar"], "lexeme":"roubado"}], "TAG":"VP"}], "
tokens":[] , "syntacticChunks":[] , "text":"cara acabou de ser roubado"}}
7 {"sentences":[{"chunks":[{"tokens":[{"features":"F=S", "POSTag":"n", "lemmas
":["cara"], "lexeme":"cara"}], "TAG":"NP"}], {"tokens":[{"features":"PR=3S=
IND", "POSTag":"v-fin", "lemmas":["acabar"], "lexeme":"acaba"}, {"features
":"-", "POSTag":"prp", "lemmas":["de"], "lexeme":"de"}, {"features":"-", "
POSTag":"v-inf", "lemmas":["ser"], "lexeme":"ser"}, {"features":"M=S", "
POSTag":"v-pcp", "lemmas":["roubar"], "lexeme":"roubado"}], "TAG":"VP"}], "
tokens":[] , "syntacticChunks":[] , "text":"cara acaba de ser roubado"}}
8 {"sentences":[{"chunks":[{"tokens":[{"features":"F=S", "POSTag":"n", "lemmas
":["cara"], "lexeme":"cara"}], "TAG":"NP"}], {"tokens":[{"features":"PS=3S=
IND", "POSTag":"v-fin", "lemmas":["ir", "ser"], "lexeme":"foi"}, {"features
":"M=S", "POSTag":"v-pcp", "lemmas":["roubar"], "lexeme":"roubado"}], "TAG
":"VP"}], {"tokens":[] , "syntacticChunks":[] , "text":"cara foi roubado"}}
9 {"sentences":[{"chunks":[{"tokens":[{"features":"M=S", "POSTag":"art", "
lemmas":["um"], "lexeme":"um"}, {"features":"M=S", "POSTag":"n", "lemmas":["
senhor"], "lexeme":"senhor"}], "TAG":"NP"}], {"tokens":[{"features":"PR=3S=
IND", "POSTag":"v-fin", "lemmas":["acabar"], "lexeme":"acaba"}, {"features
":"-", "POSTag":"prp", "lemmas":["de"], "lexeme":"de"}, {"features":"-", "
POSTag":"v-inf", "lemmas":["ser"], "lexeme":"ser"}, {"features":"M=S", "
POSTag":"v-pcp", "lemmas":["roubar"], "lexeme":"roubado"}], "TAG":"VP"}], "
tokens":[] , "syntacticChunks":[] , "text":"um senhor acaba de ser roubado
"}}}
10 {"sentences":[{"chunks":[{"tokens":[{"features":"F=S", "POSTag":"n", "lemmas
":["pessoa"], "lexeme":"pessoa"}], "TAG":"NP"}], {"tokens":[{"features":"PS
=3S=IND", "POSTag":"v-fin", "lemmas":["ir", "ser"], "lexeme":"foi"}, {"
features":"F=S", "POSTag":"v-pcp", "lemmas":["roubar"], "lexeme":"roubada
"}], "TAG":"VP"}], {"tokens":[] , "syntacticChunks":[] , "text":"pessoa foi
roubada "}}
11 {"sentences":[{"chunks":[{"tokens":[{"features":"F=S", "POSTag":"n", "lemmas
":["cara"], "lexeme":"cara"}], "TAG":"NP"}], {"tokens":[{"features":"PS=3S=
IND", "POSTag":"v-fin", "lemmas":["ir", "ser"], "lexeme":"foi"}, {"features
":"M=S", "POSTag":"v-pcp", "lemmas":["roubar"], "lexeme":"roubado"}], "TAG
":"VP"}], {"tokens":[] , "syntacticChunks":[] , "text":"cara foi roubado"}}
12 {"sentences":[{"chunks":[{"tokens":[{"features":"M=S", "POSTag":"n", "lemmas
":["homem"], "lexeme":"homem"}], "TAG":"NP"}], {"tokens":[{"features":"PS=3S
=IND", "POSTag":"v-fin", "lemmas":["ir", "ser"], "lexeme":"foi"}, {"features
":"M=S", "POSTag":"v-pcp", "lemmas":["roubar"], "lexeme":"roubado"}], "TAG
":"VP"}], {"tokens":[] , "syntacticChunks":[] , "text":"homem foi roubado"}}

```

```

13 {"sentences":[{"chunks":[{"tokens":[{"features":"F=S","POSTag":"n","lemmas":["mulher"],"lexeme":"mulher"}],"TAG":"NP"},{"tokens":[{"features":"PS=3S=IND","POSTag":"v-fin","lemmas":["ir","ser"],"lexeme":"foi"},{"features":"F=S","POSTag":"v-pcp","lemmas":["roubar"],"lexeme":"roubada"}],"TAG":"VP"}],"tokens":[],"syntacticChunks":[],"text":"mulher foi roubada"}}}
14 {"sentences":[{"chunks":[{"tokens":[{"features":"M=S","POSTag":"n","lemmas":["homem"],"lexeme":"homem"}],"TAG":"NP"},{"tokens":[{"features":"PS=3S=IND","POSTag":"v-fin","lemmas":["ir","ser"],"lexeme":"foi"},{"features":"M=S","POSTag":"v-pcp","lemmas":["assaltar"],"lexeme":"assaltado"}],"TAG":"VP"}],"tokens":[],"syntacticChunks":[],"text":"homem foi assaltado"}}}
15 {"sentences":[{"chunks":[{"tokens":[{"features":"F=S","POSTag":"n","lemmas":["mulher"],"lexeme":"mulher"}],"TAG":"NP"},{"tokens":[{"features":"PS=3S=IND","POSTag":"v-fin","lemmas":["ir","ser"],"lexeme":"foi"},{"features":"F=S","POSTag":"v-pcp","lemmas":["assaltar"],"lexeme":"assaltada"}],"TAG":"VP"}],"tokens":[],"syntacticChunks":[],"text":"mulher foi assaltada"}}}
16 {"sentences":[{"chunks":[{"tokens":[{"features":"M=S","POSTag":"n","lemmas":["refem"],"lexeme":"refem"}],"TAG":"NP"}],"tokens":[],"syntacticChunks":[],"text":"refem"}}}
17 {"sentences":[{"chunks":[{"tokens":[{"features":"F=S","POSTag":"n","lemmas":["pessoa"],"lexeme":"pessoa"}],"TAG":"NP"},{"tokens":[{"features":"PR=3S=IND","POSTag":"v-fin","lemmas":["acabar"],"lexeme":"acaba"},{"features":"-","POSTag":"prp","lemmas":["de"],"lexeme":"de"},{"features":"-","POSTag":"v-inf","lemmas":["ser"],"lexeme":"ser"}],"TAG":"VP"},{"tokens":[{"features":"F=S","POSTag":"n","lemmas":["morto"],"lexeme":"morta"}],"TAG":"NP"}],"tokens":[],"syntacticChunks":[],"text":"pessoa acaba de ser morta"}}}
18 {"sentences":[{"chunks":[{"tokens":[{"features":"M=S","POSTag":"art","lemmas":["um"],"lexeme":"um"},{"features":"M=S","POSTag":"n","lemmas":["cara"],"lexeme":"cara"}],"TAG":"NP"},{"tokens":[{"features":"PR=3S=IND","POSTag":"v-fin","lemmas":["acabar"],"lexeme":"acaba"},{"features":"-","POSTag":"prp","lemmas":["de"],"lexeme":"de"},{"features":"-","POSTag":"v-inf","lemmas":["ser"],"lexeme":"ser"},{"features":"M=S","POSTag":"v-pcp","lemmas":["roubar"],"lexeme":"roubado"}],"TAG":"VP"}],"tokens":[],"syntacticChunks":[],"text":"um cara acaba de ser roubado"}}}
19 {"sentences":[{"chunks":[{"tokens":[{"features":"M=S","POSTag":"art","lemmas":["um"],"lexeme":"um"},{"features":"M=S","POSTag":"n","lemmas":["cara"],"lexeme":"cara"}],"TAG":"NP"},{"tokens":[{"features":"PR=3S=IND","POSTag":"v-fin","lemmas":["acabar"],"lexeme":"acaba"}],"TAG":"VP"},{"tokens":[{"features":"-","POSTag":"adv","lemmas":[],"lexeme":"foi"}],"TAG":"ADVP"}],"tokens":[],"syntacticChunks":[],"text":"um cara acaba foi roubado"}}}
20 {"sentences":[{"chunks":[{"tokens":[{"features":"F=S","POSTag":"art","lemmas":["um"],"lexeme":"uma"},{"features":"F=S","POSTag":"n","lemmas":["pessoa"],"lexeme":"pessoa"}],"TAG":"NP"},{"tokens":[{"features":"PS=3S=IND","POSTag":"v-fin","lemmas":["acabar"],"lexeme":"acaba"}],"TAG":"VP"}],"tokens":[],"syntacticChunks":[],"text":"uma pessoa acaba de ser morta"}}}

```

```

=3S=IND", "POSTag": "v-fin", "lemmas": ["ir", "ser"], "lexeme": "foi"}, {"
features": "F=S", "POSTag": "v-pp", "lemmas": ["roubar"], "lexeme": "roubada
"}], "TAG": "VP"}], "tokens": [], "syntacticChunks": [], "text": "uma pessoa foi
roubada "}}
21 {"sentences": [{"chunks": [{"tokens": [{"features": "F=S", "POSTag": "art", "
lemmas": ["um"], "lexeme": "uma"}, {"features": "F=S", "POSTag": "n", "lemmas
": ["pessoa"], "lexeme": "pessoa"}], "TAG": "NP"}, {"tokens": [{"features": "PS
=3S=IND", "POSTag": "v-fin", "lemmas": ["ir", "ser"], "lexeme": "foi"}], "TAG": "
VP"}, {"tokens": [{"features": "F=S", "POSTag": "n", "lemmas": ["morta"], "
lexeme": "morta"}], "TAG": "NP"}], "tokens": [], "syntacticChunks": [], "text": "
uma pessoa foi morta "}}

```

A.6 Variedades

```

1 {"sentences": [{"chunks": [{"tokens": [{"features": "F=S", "POSTag": "adj", "
lemmas": ["boa"], "lexeme": "boa"}, {"features": "F=S", "POSTag": "n", "lemmas
": ["noite"], "lexeme": "noite"}], "TAG": "NP"}], "tokens": [], "syntacticChunks
": [], "text": "boa noite "}}
2 {"sentences": [{"chunks": [{"tokens": [{"features": "IMPF=3S=IND", "POSTag": "v-
fin", "lemmas": ["estar"], "lexeme": "estava"}, {"features": "-", "POSTag": "v-
ger", "lemmas": ["passar"], "lexeme": "passando"}], "TAG": "VP"}], "tokens
": [], "syntacticChunks": [], "text": "estava passando "}}
3 {"sentences": [{"chunks": [{"tokens": [{"features": "PS=3S=IND", "POSTag": "v-fin
", "lemmas": ["sair"], "lexeme": "saiu"}, {"features": "-", "POSTag": "v-ger", "
lemmas": ["correr"], "lexeme": "correndo"}], "TAG": "VP"}], "tokens": [], "
syntacticChunks": [], "text": "saiu correndo "}}
4 {"sentences": [{"chunks": [{"tokens": [{"features": "-", "POSTag": "prp", "lemmas
": ["em"], "lexeme": "em"}], "TAG": "PP"}, {"tokens": [{"features": "F=S", "
POSTag": "n", "lemmas": ["direcao"], "lexeme": "direcao"}], "TAG": "NP"}], "
tokens": [], "syntacticChunks": [], "text": "em direcao "}}
5 {"sentences": [{"chunks": [{"tokens": [{"features": "M=S", "POSTag": "n", "lemmas
": ["bone"], "lexeme": "bone"}, {"features": "M=S", "POSTag": "adj", "lemmas": ["
preto"], "lexeme": "preto"}], "TAG": "NP"}], "tokens": [], "syntacticChunks
": [], "text": "bone preto "}}
6 {"sentences": [{"chunks": [{"tokens": [{"features": "F=S", "POSTag": "n", "lemmas
": ["blusa"], "lexeme": "blusa"}, {"features": "F=S", "POSTag": "adj", "lemmas
": ["vermelho"], "lexeme": "vermelha"}], "TAG": "NP"}], "tokens": [], "
syntacticChunks": [], "text": "blusa vermelha "}}
7 {"sentences": [{"chunks": [{"tokens": [{"features": "M=S", "POSTag": "n", "lemmas
": ["rosto"], "lexeme": "rosto"}, {"features": "M=S", "POSTag": "v-pp", "lemmas
": ["tampar"], "lexeme": "tampado"}], "TAG": "NP"}], "tokens": [], "
syntacticChunks": [], "text": "rosto tampado "}}
8 {"sentences": [{"chunks": [{"tokens": [{"features": "M=S", "POSTag": "adj", "
lemmas": ["bom"], "lexeme": "bom"}, {"features": "M=S", "POSTag": "n", "lemmas
": ["dia"], "lexeme": "dia"}], "TAG": "NP"}], "tokens": [], "syntacticChunks
": [], "text": "bom dia "}}

```

```
9 {"sentences":[{"chunks":[{"tokens":[{"features":"M=S","POSTag":"n","lemmas":["socorro"],"lexeme":"socorro"}],"TAG":"NP"}],"tokens":[],"syntacticChunks":[],"text":"socorro"}]}
```