

## 7

### Conclusões

Este capítulo apresenta as conclusões da Tese e descreve como os objetivos propostos foram alcançados. Nele também são discutidas as recomendações para continuação da pesquisa e alterações sugeridas para aperfeiçoamento do método proposto.

#### 7.1

##### Visão Geral dos Trabalhos Desenvolvidos

A Tese propõe métodos e ferramentas para construção e análise de Mapas de Inteligência. Um Mapa de Inteligência é a representação topológica de uma investigação criminal em formato de um grafo. Para construir este mapa, é necessária a extração de conhecimento de históricos policiais, utilizando integração de algoritmos e ferramentas sistêmicas.

Uma estrutura de radicais derivados de palavras freqüentes em históricos policiais foi desenvolvida para apoio à construção de dicionários temáticos baseados na extração de palavras-chave de textos criminais. As estruturas de palavras-chave resultantes foram usadas para extração de entidades úteis de históricos policiais e modelagem de Mapas de Inteligência. Estudos de caso foram apresentados e Mapas de Inteligência foram gerados para demonstrar a aplicabilidade da metodologia de análise proposta nesta Tese.

O trabalho desenvolvido insere-se no contexto global como um conjunto completo de recursos para atividades de inteligência, orientados para acelerar o tempo da descoberta de conhecimento em investigações, integrando todo o ciclo de tratamento de informações criminais, do histórico aos registros conclusivos. Métodos manuais e ferramentas automatizadas têm sido largamente empregadas em departamentos policiais para o trabalhos de análise em bases de históricos, estruturados e não estruturados. Estes recursos presentes, entretanto, não abrangem, operando como uma ferramenta integrada, toda extensão das necessidades específicas para tratamento das informações a serem extraídas para

análise e produção de conhecimento. O conjunto de métodos, ferramentas e algoritmos inseridos neste trabalho cobrem tal extensão de necessidades para tratamento e análise de elementos criminais de utilidade extraídos diretamente de bases criminais da maioria dos formatos textuais convencionais disponíveis em meios digitais. Os principais métodos para análise criminal desenvolvidos nesta Tese foram os seguintes:

- **Método para desenvolvimento de dicionário especializado** - método para desenvolvimento de um dicionário temático especializado, a partir de radicais treinados, contendo vocabulário específico da cultura policial, destinado à extração de entidades úteis em ocorrências policiais textuais.
- **Método para classificação de subgrupos criminais**- análise de configurações organizacionais de quadrilhas criminais, usando técnicas de classificação não supervisionada de dados, através de uma ferramenta que utiliza princípios aplicados de redes neurais
- **Método para identificação de Densidade entre subgrupos criminais** - pesquisa de interseções existentes entre subgrupos criminais identificando elos comuns intermediários.
- **Método para identificação de Centralidades em subgrupos criminais** - cálculo de aproximações entre entidades entidades de um agrupamento selecionado no mapa de inteligência.
- **Método para redução / ampliação das buscas por entidades existentes nos textos** - nível variável de aprofundamento dos detalhes extraídos da base de históricos pesquisados, permitindo a ampliação ou redução comandada do tratamento de entidades extraídas. O método orienta-se por parâmetros de análise para alternativas de modelagem do mapa de Inteligência, selecionando-se opções de classificação das entidades no dicionário especialista (extração elementar, extração moderada ou extração profunda).

## 7.2

### Resumo dos Objetivos Propostos e Alcançados

Na Tese foram propostos métodos para construção de dicionários temáticos, destacando a construção de um dicionário especialista baseado na estrutura de radicais treinados e na região RV de uma palavra. O método baseado na região RV foi comparado com outros métodos pesquisados (região e regras), cujo resultado apresentou desempenho compatível com os trabalhos de Orengo e Huyck (2001). O Dicionário Especialista foi construído com vocábulos específicos da cultura policial brasileira, sendo destinado à extração de entidades úteis em históricos policiais textuais e modelagem de uma rede semântica, onde as entidades são representadas pelos vértices de um grafo dirigido e os arcos representam as relações entre as entidades. Os relacionamentos entre as entidades do grafo foram calculados aplicando-se algoritmos para cálculos de co-ocorrências. A rede desenvolvida foi utilizada para construção de um Mapa de Inteligência.

Para extração de conhecimento dos Mapas de Inteligência foram utilizados instrumentos construídos com apoio de algoritmos para tratamento de grafos e índices de centralidade em redes sociais. Dentre as ferramentas desenvolvidas para análise destacam-se a identificação das mais fortes relações entre entidades, identificação e análise de organizações criminais, pesquisa de interseções entre entidades e cálculo de centralidades em subgrupos criminais.

Três estudos de casos foram apresentados para demonstração da metodologia proposta, baseados em Mapas de Inteligência extraídos da base de ocorrências do sistema criminal do Rio de Janeiro.

No primeiro estudo construímos de um cenário envolvendo grupos de narcotraficantes atuantes na região do Rio de Janeiro. Como resultado deste estudo foi produzido um mapeamento de quadrilhas e suas respectivas organizações. Foi produzida uma análise sobre armas e explosivos utilizados e um estudo de localização dos armamentos em logradouros.

O segundo estudo tratou do assassinato do jornalista Tim Lopes. Neste estudo foram deduzidas a dinâmica do assassinato, os vínculos de autoria do crime, a organização da quadrilha do principal suspeito e seu alcance territorial.

O terceiro estudo revelou cúmplices de um seqüestro, cujos históricos não relacionavam diretamente suspeitas envolvidas em ocorrências policiais independentes. Foram reveladas relações de cumplicidade e possível origem do seqüestro através de vínculos extraídos do Mapa de Inteligência.

Apresentamos através dos resultados oferecidos nos exemplos que o método proposto nesta Tese está capacitado para interpretar massas de ocorrências policiais, extraindo conhecimento investigativo. Demonstramos como o método pode apoiar a descoberta de quadrilhas em bases criminais e auxiliar a identificação de autorias de delitos com grande economia de tempo.

Foi demonstrado que o método proposto pode tratar documentos textuais com múltiplos formatos de leitura através de um processo habilitado, após a fase de treinamento, a um trabalho com mínima interferência manual, aplicando-se dicionários especializados para suporte à transferência de conteúdos dos textos para Mapas de Inteligência.

### 7.3

#### **Contribuições oferecidas através da Pesquisa Desenvolvida**

A motivação maior para desenvolvimento da proposta oferecida nesta Tese decorre da grande incidência de crimes sem solução de autoria e do grande volume de ocorrências registrados nas grandes metrópoles brasileiras, em particular na cidade do Rio de Janeiro.

Lidando com a massa de históricos criminais somada aos registros de Inteligência e demais fontes utilizadas para investigação, a atividade policial carece de instrumentos de inteligência que auxiliem a elucidação de crimes e descoberta de conhecimento em bases de ocorrências.

O esforço policial compartilhado entre a ação do policiamento preventivo e o policiamento investigativo representa um cenário complexo para decisões do planejamento operacional. Prevenir e Investigar lidam com as mesmas variáveis representadas pelo efetivo policial que devem ser compartilhadas como recursos cooperados, mas concorrentes na busca pelos resultados finais.

A eficiência e a eficácia da Investigação policial necessita o emprego de ferramentas automatizadas que abranjam todo o ciclo da investigação, que

compreende o registro da ocorrência, análise e extração do conhecimento do histórico policial.

A principal contribuição desta Tese reside na proposição de um conjunto de métodos e ferramentas que consolidam todo o ciclo do tratamento investigativo de históricos policiais. Estes métodos abrangem a extração de informações de ocorrências policiais, a preparação de Mapas de Inteligência e a extração do conhecimento.

Fundamentado na revisão bibliográfica, não foi encontrada outra metodologia que abranja a integração total das etapas e funções relacionadas para o ciclo de investigação criminal.

O desenvolvimento de sistemas que implementam os métodos propostos e automatizam o processo de Investigação policial pode ser destacado como outra contribuição da Tese, dentre estes: algoritmo para remoção de ambigüidades encontradas em entidades selecionadas para construção de dicionários, algoritmo para construção de mapas de inteligência baseado em dicionários temáticos, algoritmo para cálculo dos vínculos entre entidades extraídas de textos e modelagem de uma rede semântica, algoritmo para modelagem de estruturas organizacionais em mapas de inteligência, algoritmo para classificação de classes organizacionais aplicando sistema de aprendizado não-supervisionado SOM, algoritmo para cálculo de centralidades em subredes criminais extraídas de mapas de inteligência. Estes sistemas melhoraram a qualidade e a velocidade do processo de elucidação de crimes, permitindo a identificação de relacionamentos ocultos entre membros de quadrilhas, locais e objetos de crime. Estes relacionamentos são evidenciados no cálculo das co-ocorrências, outro ponto inovador da tese. A totalidade da programação para os algoritmos desenvolvidos é inédita, sendo utilizada linguagem Visual Basic 6, em ambiente Windows.

Esta tese também contribui no estado-da-arte de extração de radicais para geração de palavras-chaves, pois propõe novo algoritmo para este fim. Este algoritmo apresenta bom desempenho, em termo de velocidade de extração e pertinência das informações extraídas, quando comparado com os algoritmos encontrados na literatura.

Finalmente merece ser destacado o fato de todos os sistemas serem desenvolvidos em língua portuguesa, enquanto os casos apresentados na literatura são em inglês.

## 7.4

### Recomendações para trabalhos Futuros

As recomendações seguintes são apresentadas para continuidade das pesquisas desenvolvidas nesta Tese

#### 7.4.1

##### Ampliação da capacidade de extração em textos livres

O Mapa de Inteligência deve apresentar todas as possíveis conexões entre os atores potencialmente envolvidos no delito analisado e que possam ter, direta ou indiretamente, participado de atividades no cenário representado.

As pesquisas de autoria em uma análise criminal requerem aprofundamento nas informações extraídas de textos sob formato de entidades. O aumento da capacidade de extrair informações úteis dos textos tratados aprofunda a visão dos algoritmos extratores, aumentando a eficiência na análise criminal para descoberta de linhas de vínculos “invisíveis” entre as peças que integram um cenário de uma investigação criminal.

As seguintes recomendações específicas relacionam-se com a ampliação de entidades nos dicionários temáticos. O nível de detalhamento dos Dicionários de Logradouros deve ser ampliado para incorporar novas referências geográficas, como distritos, vilas, favelas e outras referências importantes e usuais. O Dicionário Especialista necessita maior abrangência de siglas, marcas de veículos, especificidades de armas, munições, referências cronológicas, numéricas e quantificações. Importante citar que a inclusão de novas palavras ou referências no dicionário especialista representa um alto investimento em treinamento das estruturas resultantes para torná-las adequadas e eficientes como ferramenta de apoio à novas extrações de entidades.

#### 7.4.2

##### Identificação de nomes próprios

Um problema foi identificado com as regras para extração de nomes próprios em textos parcialmente ou totalmente formatados em maiúsculas.

Entidades constituídas por nomes próprios são regularmente extraídas de forma automática, por meio de regras sintáticas que identificam as letras maiúsculas nos textos pesquisados. Frequentemente foram encontrados nos históricos criminais tratados por esta Tese textos formatados parcialmente ou totalmente em maiúsculas, o que torna inútil a regra para extração automática dos nomes próprios identificados. Procedimentos manuais para captação de entidades não extraídas automaticamente precisam ser executados para evitar a perda de entidades de importância para investigação.

Recomendamos a inclusão de regras complementares para otimizar a identificação automática de nomes próprios a partir de textos formatados em maiúsculas permitindo a extração de pessoas, logradouros, organizações e marcas, através da análise do texto. Uma alternativa possível para identificação de nomes próprios permite a análise de antecedentes referenciais como títulos e tipos de logradouros, como “**DR. BOAMORTE**”, “**RUA ADALBERTO ARANHA**”, “**CAMINHONETE FORD**”, “**FAVELA DA BARRA FUNDA**”, “**VILA DO PANAMERICANO**”, ou “**PISTOLA P-40**”.

### 7.4.3

#### **Algoritmos para redução do tempo da Extração**

Em decorrência do volume potencial das inclusões propostas para as novas categorias de entidades a serem incorporadas nos dicionários temáticos, o consumo de tempo necessário para extração de entidades será elevado devido à complexidade dos algoritmos envolvidos no processamento e do tempo para acesso às estruturas dos dicionários temáticos. Como alternativa para minimizar o tempo de processamento da extração recomendamos a incorporação de algoritmo para otimizar o acesso às estruturas dos dicionários temáticos.

Uma possível alternativa foi aplicada na otimização do acesso à estrutura de stop-words, intensamente utilizada durante a extração de radicais e palavras chave para preparação do dicionário especialista, aplicando uma lista binária *hash-code* para bloqueio de palavras indesejáveis.

#### 7.4.4

#### **Testes exaustivos para cálculos de Co-Ocorrências**

Co-ocorrências é a base da heurística aplicada para cálculo do valor do relacionamento entre entidades extraídas de textos para modelagem de Mapas de Inteligência. O Fator de Importância da Entidade na Extração é um parâmetro de relevância usado como fator de ponderação do tipo da entidade na extração aplicado para cálculo das co-ocorrências, podendo variar de acordo com os propósitos de cada investigação desenvolvida. Os estudos para cálculos das co-ocorrências não foram exaustivos. Ensaio numéricos devem ser conduzidos para melhor ajuste do Fator de Importância da Entidade na Extração, buscando a determinação de valores mais precisos e adequados para os tipos de cenários modelados em uma investigação.

Devido ao alto consumo de recursos computacionais envolvidos nos cálculos das co-ocorrências, particularmente quando são utilizadas grandes massas documentais, recomendamos que os estudos para redução de tempo de processamento, conforme sugerido no tópico 7.4.3 (otimização do acesso aos dicionários temáticos) devam preceder os estudos para aperfeiçoamento dos cálculos de co-ocorrências.