

UNIVERSIDADE REGIONAL DE BLUMENAU
CENTRO DE CIÊNCIAS EXATAS E NATURAIS
CURSO DE CIÊNCIA DA COMPUTAÇÃO – BACHARELADO

GEO-OP: SISTEMA DE MAPEAMENTO GEOGRÁFICO DE
OCORRÊNCIAS POLICIAIS

ALEX AUGUST SCHLOTE

BLUMENAU
2016

2016/1

ALEX AUGUST SCHLOTE

**GEO-OP: SISTEMA DE MAPEAMENTO GEOGRÁFICO DE
OCORRÊNCIAS POLICIAIS**

Trabalho de Conclusão de Curso apresentado ao curso de graduação em Ciência da Computação do Centro de Ciências Exatas e Naturais da Universidade Regional de Blumenau como requisito parcial para a obtenção do grau de Bacharel em Ciência da Computação.

Prof. Daniel Theisges dos Santos, Mestre - Orientador

**BLUMENAU
2016**

2016/1

GEO-OP: SISTEMA DE MAPEAMENTO GEOGRÁFICO DE OCORRÊNCIAS POLICIAIS

Por

ALEX AUGUST SCHLOTE

Trabalho de Conclusão de Curso aprovado
para obtenção dos créditos na disciplina de
Trabalho de Conclusão de Curso II pela banca
examinadora formada por:

Presidente: _____
Prof. Daniel Theisges dos Santos, Mestre – Orientador, FURB

Membro: _____
Prof. Dalton Solano dos Reis , Mestre – FURB

Membro: _____
Prof. Alexander Roberto Valdameri, Mestre – FURB

Blumenau, dia 4 de Julho de 2016

Dedico este trabalho aos meus amigos e familiares.

AGRADECIMENTOS

Dedico esse trabalho as minhas irmãs Kysa Schlote e Kelly Schlote e também minha mãe Itamici Eleutério de Souza, por todo o auxílio e por sempre estarem ao meu lado nos momentos difíceis.

Aos meus amigos do Bruf, por toda paciência, ajuda e compartilhamento de conhecimento, no qual tornaram esses anos da graduação inesquecíveis.

Agradeço aos senhores Atanásio Antunes Vieira coordenador do sistema SISP e Alfeu Orben delegado da Polícia Civil e gestor do sistema SISP pela disponibilização dos dados de relatórios policiais, no qual foi possível o desenvolvimento desse trabalho.

Ao meu orientador Daniel Theisges dos Santos, pelos conselhos, empenho e garra durante o desenvolvimento desse trabalho.

Se não gostam de você, você tem que estar sempre certo, ou não vale nada

Dr. Gregory House

RESUMO

Este trabalho apresenta o desenvolvimento de um sistema *web* de mapeamento geográfico de ocorrências policiais da cidade de Blumenau. Esses tipos de sistemas são fundamentais para a análise criminal de uma cidade, onde é possível acompanhar o planejamento do policiamento através das ocorrências de uma determinada localidade. As ocorrências são buscadas através de páginas de pesquisas de notícias e de relatórios policiais. Visto que as ocorrências não se encontram estruturadas, técnicas de mineração de texto são utilizadas através da abordagem de análise semântica, com o intuito de classificar e estruturar os dados de ocorrências conforme sua natureza. As ocorrências coletadas são apresentadas em um mapa de informação geográfica *web*, sendo possível visualizar as ocorrências através de mapas de calor e de mapas de pontos, permitindo também ao usuário filtrar as ocorrências por classificação, data início e data fim. O sistema foi desenvolvido em Java, utilizando bibliotecas terceiras para coleta de texto nas páginas *web*, busca da latitude e longitude e para execução das etapas de mineração de texto sobre as ocorrências policiais. A *API* em JavaScript do Google Maps foi utilizada para a geolocalização das ocorrências e a criação do mapa de pontos e de calor. Com o resultado do desenvolvimento desse sistema, é possível observar os locais da cidade onde ocorrem maiores crimes, como furtos, assaltos, roubos e etc.

Palavras-chave: Violência. Mineração de texto. SIG. Mapas de calor.

ABSTRACT

This work presents the development of a geographical mapping system web of police occurrences from the city of Blumenau. These types of systems are fundamental to the criminal analysis of a city, where you can follow the planning of policing through the occurrences of a particular locality. The occurrences are sought through news search pages and police reports. The occurrences are not structured text-mining techniques are used through the semantic analysis approach, in order to classify and structure instances data according to their nature. Collected events are displayed on a map web geographic information, and you can view the events through heat maps and points of maps, also allowing the user to filter the events by rating, date, start and end date. The system was developed in Java, using the 3rd libraries for text collection in web pages, search latitude and longitude and implementation of text mining steps on the police occurrences. The Google Maps JavaScript API was used for geolocation of events and creation of map points and heat. With the result of the development of this system, it is possible to see the places of the city where there are major crimes, such as thefts, burglaries, robberies, etc.

Key-words: Violence. Text Mining. GIS. Heat MAPS.

LISTA DE FIGURAS

Figura 1 - <i>Online Crime Maps</i>	18
Figura 2 - Tipos de descoberta de conhecimento	19
Figura 3 - Etapas do KDD	20
Figura 4 - Técnicas de classificação de dados	20
Figura 5 - Etapas da Mineração de Texto.....	23
Figura 6 - Tela de aplicação do GMAPS.....	24
Figura 7 - Arquitetura do GME	25
Figura 8 - Exemplo de mapa de calor pela ferramenta GMAPS	26
Figura 9 - Mapa criminal de Blumenau	27
Figura 10 - Mapa de calor WikiCrimes	29
Figura 11 - Mapa de homicídios cidade de Londrina.....	30
Figura 12 - Diagrama de caso de uso	32
Figura 13 - Diagrama de pacotes	34
Figura 14 - Diagrama de classes do Schedule	35
Figura 15 - Diagrama de classe do GEO-OP.....	36
Figura 16 - Relatórios de atividades policias sala de notícias	37
Figura 17 - Página web de ocorrências policias	38
Figura 18 - Relatório de ocorrências policias extensão .csv	44
Figura 19 - Tela inicial do sistema	46
Figura 20 - Legendas do mapa de pontos do <i>site</i> de notícias	47
Figura 21 - Mapa de pontos do sistema.....	48
Figura 22 - Mapa de calor do sistema.....	48
Figura 23 - Ocorrências em Blumenau no mês de janeiro de 2016 com mapa de pontos.....	52
Figura 24 - Furtos de Blumenau ano de 2015 na região central com o mapa de calor	53
Figura 25 - Furtos de Blumenau ano de 2015 na região norte com o mapa de calor	53
Figura 26 - Ameaças em Blumenau ano de 2015 com o mapa de calor.....	54
Figura 27 - Diagrama classe Schedule com atributos e métodos	60

LISTA DE QUADROS

Quadro 1 - Coleta de dados dos relatórios de atividades policias no <i>site</i> sala de notícias	38
Quadro 2 - Coleta de todos os <i>links</i> de ocorrências.....	39
Quadro 3 - Coleta das ocorrências pelo <i>site</i>	39
Quadro 4 - Exemplo de ocorrências policiais do <i>site</i>	40
Quadro 5 - Busca da latitude e longitude da ocorrência.....	40
Quadro 6 - Pré-processamento das ocorrências policias do <i>site</i> Sala de Notícias.....	41
Quadro 7 - Exemplo de ocorrência policial do <i>site</i> Sala de Notícias	42
Quadro 8 - Classificação da ocorrência sala de notícias	43
Quadro 9 - Classificação das ocorrências de relatórios policiais	44
Quadro 10 - Requisição via Servlet do Servidor	45
Quadro 11 - Criação de mapa de calor	45
Quadro 12 - Ocorrências não classificadas do <i>site</i>	50
Quadro 13 - Classificação da ocorrência errada do <i>site</i>	50
Quadro 14 - Comparação de trabalhos correlatos e sistema desenvolvido	54
Quadro 15 - Lista de <i>StopWords</i>	61
Quadro 17 - Lista de classificações das ocorrências das páginas de notícias.....	62
Quadro 18 - Lista de classificações das ocorrências dos relatórios policiais	62
Quadro 19 - Verbos para classificações	63
Quadro 20 - Crimes policiais para classificação páginas de notícias	63
Quadro 21 - Objetos policiais para classificações	63
Quadro 22 - Crimes policiais para classificação dos relatórios policiais	63
Quadro 23 - Crimes policiais que precisam de objeto policial.....	63

LISTA DE TABELAS

Tabela 1 – Classificações das ocorrências realizadas pelo algoritmo no *site*.....49

Tabela 2 - Classificações das ocorrências realizadas pelo algoritmo nos relatórios da polícia51

LISTA DE ABREVIATURAS E SIGLAS

API – *Application Programming Interface*

BBC – *British Broadcasting Corporation*

BOU – Boletim de Ocorrência Unificado

DT – *Data Warehouse*

GME – *Google Maps Engine*

GMAPS – Google Maps

HTML – *HyperText Markup Language*

KDD – *Knowledge Discovery in Databases*

KDT – *Knowledge Discovery from Text*

JSON – *JavaScript Object Notation*

MT – Mineração de Texto

PCSC – Polícia Civil de Santa Catarina

PMB – Polícia Militar de Blumenau

RF – Requisitos Funcionais

RNF – Requisitos Não Funcionais

SIG – Sistema de Informação Geográfica

SISP – Sistema Integrado de Segurança Pública

TIC – Tecnologia da Informação e da Comunicação

SUMÁRIO

1 INTRODUÇÃO.....	14
1.1 OBJETIVOS.....	15
1.2 ESTRUTURA.....	15
2 FUNDAMENTAÇÃO TEÓRICA	16
2.1 VIOLÊNCIA	16
2.1.1 Uso da informática para combater a violência.....	17
2.1.2 Uso de mapas interativos para informações de violência	17
2.2 DESCOBERTA DE CONHECIMENTO.....	18
2.2.1 Descoberta de Conhecimento em Dados Estruturado	19
2.2.2 Descoberta de Conhecimento em Dados não Estruturados.....	21
2.2.3 Mineração de Texto.....	21
2.3 SIG – SISTEMA DE INFORMAÇÃO GEOGRÁFICA	23
2.3.1 Google Maps	24
2.3.2 Mapas de calor	25
2.4 TRABALHOS CORRELATOS	26
2.4.1 Mapeamento criminal por meio da plataforma Google Maps	27
2.4.2 WikiCrimes – Um Sistema Colaborativo para Mapeamento Criminal.....	28
2.4.3 O uso do geoprocessamento na segurança pública do Estado do Paraná	29
3 DESENVOLVIMENTO	31
3.1 REQUISITOS.....	31
3.2 ESPECIFICAÇÃO	32
3.2.1 Diagrama de casos de uso	32
3.2.2 Diagrama de pacotes	33
3.2.3 Diagrama de classes	34
3.3 IMPLEMENTAÇÃO	36
3.3.1 Técnicas e ferramentas utilizadas.....	37
3.3.2 Operacionalidade da implementação	46
3.4 RESULTADOS E DISCUSSÕES.....	48
3.4.1 Resultados das coletas e classificações dos dados nas páginas de notícias	49
3.4.2 Resultados das coletas e classificações dos dados dos relatórios policiais.....	51
3.4.3 Regiões com maiores índices de violências	52

3.4.4 Comparação de trabalhos correlatos e o sistema desenvolvido	54
4 CONCLUSÕES.....	55
4.1 EXTENSÕES	56
APÊNDICE A – DIAGRAMA DA CLASSE SCHEDULE COM ATRIBUTOS E MÉTODOS.....	60
APÊNDICE B – LISTA DE <i>STOPWORDS</i> REMOVIDA PELO ALGORITMO	61
APÊNDICE C – LISTA DE TIPOS DE CLASSIFICAÇÕES DAS OCORRÊNCIAS ...	62
APÊNDICE D – LISTAS DE PALAVRAS PARA CLASSIFICAÇÕES DAS OCORRÊNCIAS	63

1 INTRODUÇÃO

O aumento da violência é um fenômeno preocupante, cujo o principal motivo está relacionado com a crescente urbanização acelerada e desigual social nas grandes cidades. Para contornar este fenômeno, países do Norte do globo decidiram controlar o crescimento da urbanização, entretanto países do Sul do globo, como o Brasil, não conseguiram controlar tal crescimento (SANTANA; SILVA; CHAGAS, 2013).

Segundo Santana, Silva e Chagas (2013, p. 1626), “o crescimento da urbanização trouxe como consequência a marginalização dos centros urbanos, prostituição, tráfico de drogas, aumento da violência, entre outros”. A desigualdade social e econômica também se manifestou com o crescimento da urbanização, o que também impactou no crescimento da violência urbana (FERREIRA; PENNA, 2005). Estudos sobre a violência urbana apresentam índices de desigualdade econômica são preponderantes dentre os fatores associados à criminalidade violenta. Os resultados desses estudos mostram também que certas áreas de desigualdade social e criminalidade violenta são mais elevadas. A estrutura populacional, onde áreas densamente povoadas tendem a ser mais acometidas por crimes violentos (BEATO, 1998).

Segundo Ferreira e Penna (2005, p. 167), “a utilização de análise de áreas geográficas sobre a violência, permite mostrar uma outra dimensão da criminalidade”. O processo de criações de sistemas de produção e reprodução da violência sobre um espaço urbano, pode-se considerar um valor estratégico ao combate da criminalidade.

A utilização de uma ferramenta para fazer análise de áreas geográficas torna-se relevante para diversas áreas de aplicação, sendo muito comum em planejamentos urbanos. Mapas de delitos fazem parte de um importante instrumento de estudos sobre a relação entre espaço e criminalidade, bem como no planejamento de ações de combate à delinquência, fato que explica sua crescente difusão. Esses mapas servem para indicar os locais prioritários para atuação mais intensa da polícia, de forma a evitar que pequenos delitos criem um ambiente propício para o surgimento de atos mais violentos (LOBO; GUIMARÃES, 2013).

A importância de realizar uma análise de áreas geográficas também depende dos dados de violência que são fornecidos pelas instituições públicas policiais. Essas informações podem ser obtidas tanto em *site* de notícias como pelas próprias instituições públicas policiais. A Lei nº 12.527 (BRASIL, 2011), conhecida também como Lei de Acesso à Informação, garante a efetividade de acesso às informações públicas, as quais podem ser obtidas gratuitamente e sem a necessidade de apresentar o motivo.

Com acesso aos dados de violência, torna-se necessário utilizar técnicas computacionais para a extração de conhecimento, principalmente quando esses dados não são estruturados, ou seja, textos manuscritos ou digitados. O crescimento de armazenamento de dados não estruturados, proporcionou o desenvolvimento das técnicas de Mineração de Texto (MT). Com essa técnica, é possível transformar dados não estruturados em dados estruturados, permitindo assim a extração de conhecimento (ARANHA; PASSOS, 2006).

Diante do exposto, foi desenvolvido um sistema capaz de coletar informações de violência através de ocorrências policiais usando MT e apresentar as suas localizações geográficas em um Sistema de Informação Geográfica (SIG), possibilitando a visualização do local da criminalidade, por data, tipos de crimes e mapas de pontos e de calor.

1.1 OBJETIVOS

Este trabalho tem como objetivo desenvolver um SIG para mapeamento geográfico de ocorrências policiais e de áreas de criminalidade, através de técnicas de mineração de texto para estruturação e classificação dos dados.

Os objetivos específicos são:

- a) minerar dados presentes no relatório de atividades operacionais da polícia militar;
- b) classificar as ocorrências encontradas de acordo com as categorias previstas no sistema;
- c) identificar a localização geográfica destas ocorrências e apresentar esses dados no SIG através de mapas de pontos e mapas de calor.

1.2 ESTRUTURA

Este trabalho está dividido em quatro capítulos. O primeiro capítulo apresenta a introdução do trabalho e os objetivos. O segundo capítulo apresenta a fundamentação teórica sobre violência, descoberta de conhecimento, MT e SIG. No terceiro capítulo é demonstrado o desenvolvimento do sistema com requisitos, especificação, implementação, resultados e operacionalidade da aplicação. Por fim, no quarto capítulo são relatadas as conclusões e também as possíveis extensões.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo são apresentados aspectos teóricos relacionados ao trabalho. A seção 2.1 expõe informações sobre a violência no Brasil e o uso de mapas e estatísticas sobre a violência. Na seção 2.2 são apresentados os principais conceitos sobre Descoberta de Conhecimento, bem como as técnicas utilizadas. A seção 2.3 aborda ferramentas e técnicas de Sistemas de Informação Geográfica.

2.1 VIOLÊNCIA

Segundo Camargo (2007), “a violência é um fenômeno histórico na constituição da sociedade brasileira”. Os fatores históricos que contribuíram para o aumento da violência no Brasil foram: a escravidão, a colonização mercantilista, o coronelismo, as oligarquias antes e depois da independência, somados a um Estado caracterizado pelo autoritarismo burocrático.

Outro fator que colaborou para aumentar a violência, está relacionado com as urbanizações aceleradas, que trazem um grande fluxo de pessoas para as áreas urbanas e assim contribuem para um crescimento desordenado e desorganizado das cidades. Um Estado ineficiente e sem programas de políticas públicas de segurança contribui para aumentar a sensação de injustiça e impunidade, que é, talvez, a principal causa da violência (CAMARGO, 2007).

Durante as últimas décadas, a violência e o risco de se tornar vítima de um crime têm sido temas frequentemente abordados em diversos estudos sociológicos e criminológicos. No Brasil, as informações mais utilizadas para o estudo da violência advêm de duas fontes oficiais: as estatísticas da polícia e da saúde. Estas fontes são claramente importantes no desenvolvimento, monitoramento e avaliação de políticas públicas, e projetam de modo coerente as taxas na área da violência (BORGES, 2013).

Informações de estudos estatísticos de violência são frequentemente noticiados. Um estudo realizado com o cruzamento dos dados da Polícia Militar de Blumenau (PMB) e do Jornal Santa Catarina mostrou que foram furtadas 450 casas em Blumenau no período de dezembro de 2013 até novembro de 2014. Os números de ocorrências no mês de janeiro de 2014 se destaca dos demais meses por causa do período de férias, onde existe cerca de 65 ocorrências registradas nesse mês (JORNAL DE SANTA CATARINA, 2016).

Apesar do panorama caótico que impera no campo da segurança pública do Brasil, a facilidade proporcionada com os avanços tecnológicos no campo da informática apresenta atualmente um avanço no campo do uso de geotecnologias. A utilização da geotecnologia pelas forças policiais influencia uma melhor utilização nos recursos humanos e materiais, no

qual é possível realizar análises de crimes em certas regiões e traçar pontos estratégicos ao combate ao crime (BORDIN et al., 2013).

2.1.1 Uso da informática para combater a violência

Depois de passar por três grandes eras – a agrícola, a industrial e a da informação –, hoje admite-se que estamos na era do conhecimento. Diante desta realidade, as polícias também têm procurado fazer uso de conhecimentos que possam ser úteis ao melhor desempenho de suas atribuições. De forma mais tardia que em outros setores, porém definitiva, a tecnologia tem se incorporado ao trabalho diário das organizações policiais (BORNHOFEN; TENFEN 2009).

Segundo Paula, Dandolini e Souza (2012, p. 120), o uso da TIC (Tecnologia da Informação e da Comunicação), oferecem uma série de facilidades para a administração e a gestão pública melhorarem os seus serviços. Um exemplo de TIC utilizado na gestão pública para combater a violência é a utilização de ferramentas como *Data Warehouse* (DT).

Segundo Santos et al (2007, p. 1915), DT é um conjunto de técnicas e bancos de dados integrados e projetados para suportar as funções dos sistemas de apoio à tomada de decisão, onde cada unidade de dados está relacionada a um determinado assunto ou fato. Com a DT é possível que os serviços de segurança pública entendam o fenômeno da criminalidade, estudando seus efeitos e consequências. (SANTOS et al, 2007).

Com os dados já prontos, é necessária uma forma para serem representados essas informações de forma interativa e atualizadas e nos últimos anos os mapas de pontos e de calor são usados para apresentar essas informações (CHANEY, 2009 apud MORICONI, 2009).

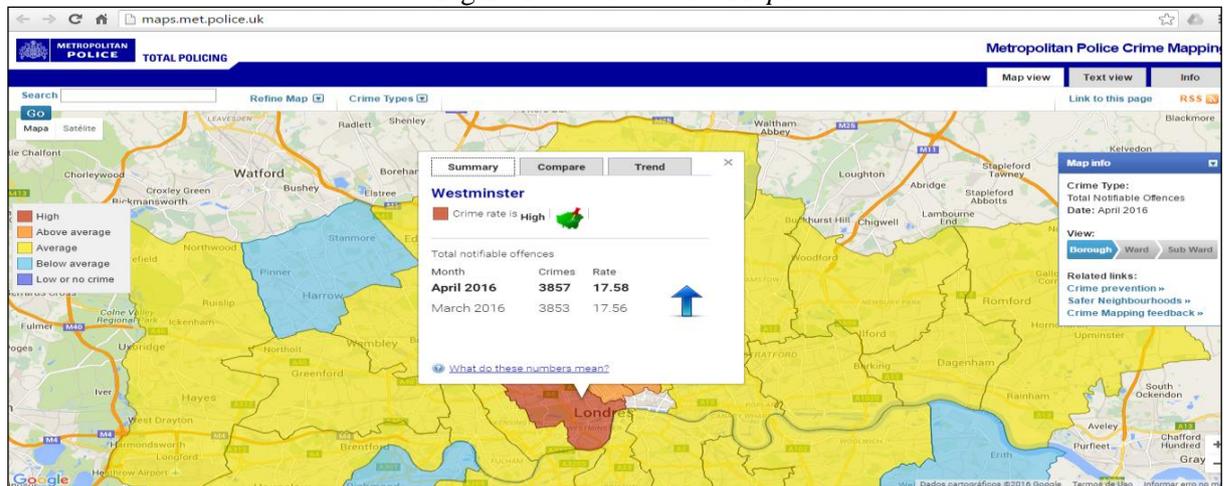
2.1.2 Uso de mapas interativos para informações de violência

Segundo Smith (2009 apud MORICONI, 2009), “ao oferecermos mapas interativos e atualizados, podemos informar melhor as pessoas sobre as áreas em que habitam, permitindo que tenham mais voz nas decisões sobre as prioridades das polícias locais”. Isso de fato cria um novo panorama a respeito da violência da região onde os habitantes moram, criando um ar mais repleto de segurança de informações sobre violência.

Moradores de países como Inglaterra e País de Gales podem atualmente conectar e navegar pelas estatísticas de crime da sua região com a ferramenta *Online Crime Maps*, que foi lançada em janeiro de 2009 (MORICONI, 2009). A ferramenta detalha graficamente informações sobre diversas ocorrências, contendo até projeções sobre as tendências de subida

e queda dos crimes. A Figura 1 demonstra a ferramenta *Online Crime Maps* da cidade de Londres, com informações do distrito Westminster no centro de Londres, onde a taxa de crime é consideravelmente alta com mais de 3 mil registros de crimes.

Figura 1 - *Online Crime Maps*



Fonte: Metropolitan Police (2016).

Para a criação de mapas interativos e atualizados, torna-se necessário obter dados através de fontes estruturadas ou não estruturadas, onde seja possível descobrir algum conhecimento que sejam tornados representativos.

2.2 DESCOBERTA DE CONHECIMENTO

O ser humano necessita de conhecimento para sobreviver, para adquirir conhecimento o homem necessita obter informações, que por sua vez são coletadas a partir da interação com outros seres e objetos. Descobrir conhecimento significa identificar, receber informações relevantes, processá-las e agregá-las ao conhecimento básico do usuário, mudando o estado de seu conhecimento atual (WIVES, 2000). Neste sentido, observa-se que o processo de descoberta de conhecimento está fortemente relacionado à forma pela qual a informação é processada.

O ser humano não consegue trabalhar com grandes volumes de informações extensas, neste sentido, mecanismos automáticos de processamento tendem a tornar o processo de descoberta de conhecimento mais eficiente. Assim, torna-se necessário automatizar o processo, utilizando softwares e computadores (MORAIS; AMBRÓSIO, 2007, p. 2).

Neste contexto surge a Descoberta de Conhecimento Apoiada por Computador (*Knowledge Discovery - KD*), que é um processo de análise de dados ou informações, cujo principal objetivo é fazer com que as pessoas possam extrair novos conhecimentos a partir da manipulação de grandes quantidades de dados.

Basicamente, existem duas abordagens utilizadas nesta área, a Descoberta de Conhecimento em Dados Estruturados e a Descoberta de Conhecimento em Dados não Estruturados. A Figura 2 apresenta uma forma de hierarquia dos tipos de descoberta de conhecimento.

Figura 2 - Tipos de descoberta de conhecimento



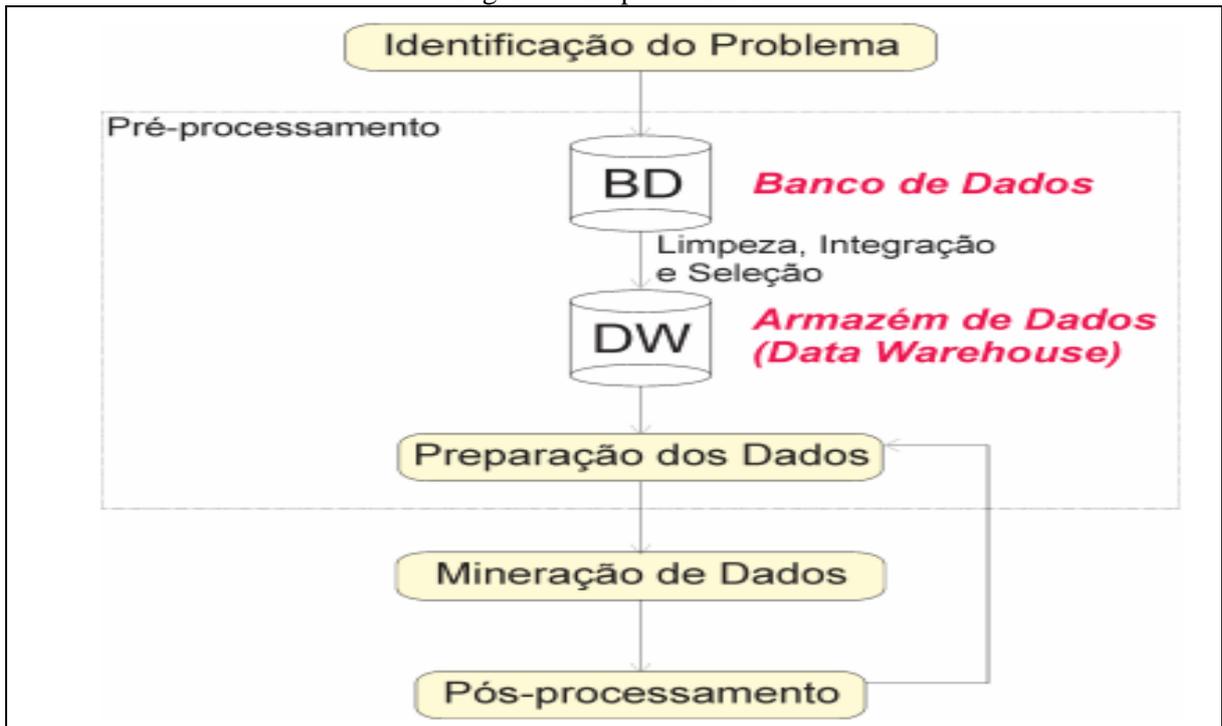
Fonte: Morais e Ambrósio (2007, p. 2).

2.2.1 Descoberta de Conhecimento em Dados Estruturado

A Descoberta de Conhecimento em Dados Estruturado está relacionada à descoberta de relacionamentos e dados implícitos em registros a base de dados. Seu principal objetivo é encontrar conhecimento em um conjunto de dados para ser utilizado em algum processo decisório. Dessa forma ela também é bastante mencionada como o Descoberta de Conhecimento em Bancos de Dados ou *Knowledge Discovery in Databases* (KDD) (MORAIS; AMBRÓSIO, 2007).

O KDD utiliza uma metodologia que segue um conjunto de fases que podem ser específicas para uma determinada aplicação ou podem ser genéricas. As principais fases do KDD são: identificação do problema, pré-processamento ou preparação dos dados, mineração de dados e pós-processamento (MORAIS; AMBRÓSIO, 2007, p. 3). A Figura 3 são demonstradas as principais fases do KDD.

Figura 3 - Etapas do KDD

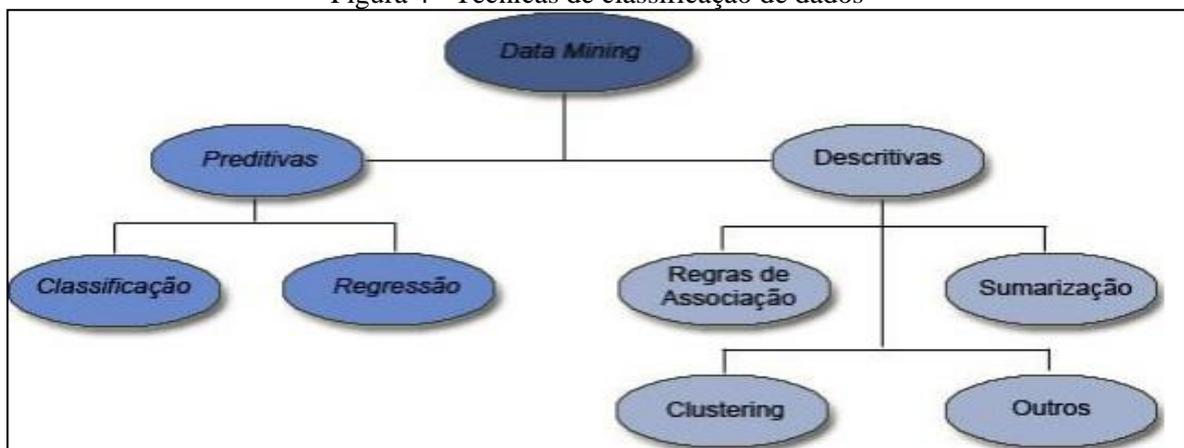


Fonte: Moraes e Ambrósio (2007, p. 3).

A mineração de dados é a fase principal do processo de KDD. Dentro dela são incluídas tecnologias de banco de dados, inteligência artificial, aprendizado de máquina, redes neurais e etc. (MORAIS; AMBRÓSIO, 2007). Nessa etapa também é feita a descoberta de conhecimento sobre os dados, no qual possui diversas tarefas que podem ser aplicadas.

Uma dessas tarefas de descoberta de conhecimento é a tarefa de Agrupamento ou *Clustering*, que visa identificar e aproximar coleções de registros similares e dividir de outros registros dissimilares. Essa etapa, não tem pretensão de classificar, estimar ou prever o valor de uma variável, apenas identificar grupos com dados similares (CAMILO; SILVA, 2009). A Figura 4 apresenta algumas das técnicas de classificação de dados.

Figura 4 - Técnicas de classificação de dados



Fonte: Rezende (2003).

2.2.2 Descoberta de Conhecimento em Dados não Estruturados

Área também conhecida como Descoberta de Conhecimento em Textos ou *Knowledge Discovery from Text* (KDT), onde seu objetivo é extrair dados desconhecidos a partir de fontes textuais (correio, imprensa, *websites*, *newsgroups*, fóruns, etc.) úteis para tomar decisões cruciais de negócio. De acordo com Beppler e Fernandes. (2005), KDT engloba técnicas e ferramentas inteligentes e automáticas que auxiliam na análise de grandes volumes de dados com o intuito de “garimpar” conhecimento útil, beneficiando qualquer domínio que utiliza textos não estruturados.

Na prática, o processo de KDT é centrado no processo de MT, que é um campo multidisciplinar, que envolve diversos processos. Na próxima seção, é apresentada informações a respeito do MT.

2.2.3 Mineração de Texto

Mineração de Textos é um conjunto de métodos usados para navegar, organizar, encontrar e descobrir informação em bases textuais. A MT surgiu a partir da necessidade de se descobrir, de forma automática, informações (padrões e anomalias) em textos. O uso dessa tecnologia permite recuperar informações, extrair dados, resumir documentos, descobrir padrões, criar associações e regras e realizar análises qualitativas ou quantitativas em documentos de texto (ARANHA; PASSOS, 2006, p. 2).

O processo de obtenção de conhecimento a partir de MT é oriundo a partir de bases de dados textuais, ou seja, documentos em linguagem natural, e que, portanto, possuem pouca ou nenhuma estrutura de dados. Além disso, os documentos escritos em linguagem natural não possuem garantia alguma de integridade (SOARES, 2013, p. 31).

Em geral, documentos em linguagem natural referem-se a massas de informação, quase sempre digitalizadas, e que não possuem rígida estrutura de dados. Exemplos deste tipo de dados são áudio, vídeo e texto livre, como por exemplo, o corpo de um e-mail (SOARES, 2013, p. 32).

Segundo Aranha (2007, p. 40), “o processo de MT contém quatro macros etapas: coleta, pré-processamento, indexação e análise da informação”. A seguir são exemplificadas cada uma das etapas.

- a) coleta: Aranha (2007, p. 40) diz que é “a etapa inicial do processo e tem como objetivo a coleta das informações que vão compor a base textual de trabalho, isto é, determinar e selecionar o universo de atuação das técnicas de mineração de texto”;
- b) pré-processamento: etapa responsável por preparar os dados para serem trabalhados

computacionalmente. É uma fase importante, uma vez que a qualidade dos resultados é diretamente influenciada pela qualidade dos dados de entrada, sendo assim, se estes não forem pré-processados adequadamente poderão prejudicar os resultados (RODRIGUES, 2014, p. 45). Dentro dessa etapa podem-se citar algumas técnicas:

- Filtering: é o processamento de remoção dos caracteres de pontuação, esses caracteres de maneira geral não alteram o significado do documento,
 - Tokenization: é o processo que converte um texto livre em unidades únicas com significado, é utilizada para decompor o documento em cada termo que o compõe,
 - StopWord: é o processo de remoção de palavras que são muito utilizadas no texto e que não trazem significado direto a uma frase,
 - Pruning: após a etapa “Stopword”, ainda existem palavras que não trazem valor ao texto, por serem usadas com muita ou pouca frequência. Portanto na etapa de pruning as palavras que aparecem com muita e pouca frequência são removidas do texto;
- c) indexação: Sabino (2014, p. 39) diz que a “indexação é a fase responsável por criar estruturas de dados denominadas índices, capazes de permitir a consulta a uma informação gerada com base na etapa de pré-processamento, sem que seja necessário analisar toda a base textual”;
- d) análise da informação: após a fase de mineração de dados, o processo entra na fase de avaliação e interpretação dos resultados. É a parte onde o usuário pode dar o julgamento final sobre a aplicabilidade dos resultados (ARANHA, 2007 p. 59). Na análise da informação, existe dois tipos de abordagens, a análise semântica e a análise estática. Abaixo é comentado sobre esses dois tipos de abordagens:
- Análise semântica: este tipo de análise emprega técnicas que avaliam a sequência dos termos no contexto dos textos, no sentido de identificar qual a sua função. Ou seja, ela procura identificar a importância das palavras dentro da estrutura de suas orações (MORAIS; AMBRÓSIO, 2007, p. 8),
 - Análise estática: esse tipo de análise emprega a importância de um termo é dado pelo número de vezes que este aparece no texto. Basicamente, seu processo envolve aprendizado estatístico a partir de dados, que normalmente inclui as etapas de codificação dos dados, estimativa dos dados e modelos de representação de documentos (MORAIS; AMBRÓSIO, 2007, p. 8);

A Figura 5 apresenta as etapas da mineração de texto, conforme foi abordado nos tópicos acima.

Figura 5 - Etapas da Mineração de Texto



Fonte: Aranha (2007).

Após realizar as etapas da MT e os dados já estarem estruturados, é possível realizar técnicas de KDD a partir deles. Também é possível, utilizar esses dados para serem representados em outras ferramentas tecnológicas, como análise de dados espacial em SIG.

2.3 SIG – SISTEMA DE INFORMAÇÃO GEOGRÁFICA

Segundo Queiroz (2009, p. 30) "a análise de dados espacial consiste em um conjunto de ferramentas e procedimentos que procura descrever, explorar padrões e avaliar a existência de relações entre um conjunto de dados obtidos a partir de um processo gerador atuando no espaço". Já Rodrigues (2014, p. 31) descreve análise espacial como "[...] uma ferramenta que possibilita manipular dados espaciais de diferentes formas e extrair conhecimento adicional como resposta [...]".

Uma das categorias de análise espacial de dados é a visualização ou apresentação dos dados. Essa etapa consiste em utilizar ferramentas como mapas e gráficos na busca de padrões e evidências sobre hipóteses pré-estabelecidas. Nos dias de hoje, são utilizados com certa simplicidade e flexibilidade softwares geográficos para a representação dessa etapa em mapas (QUEIROZ, 2009 p. 30).

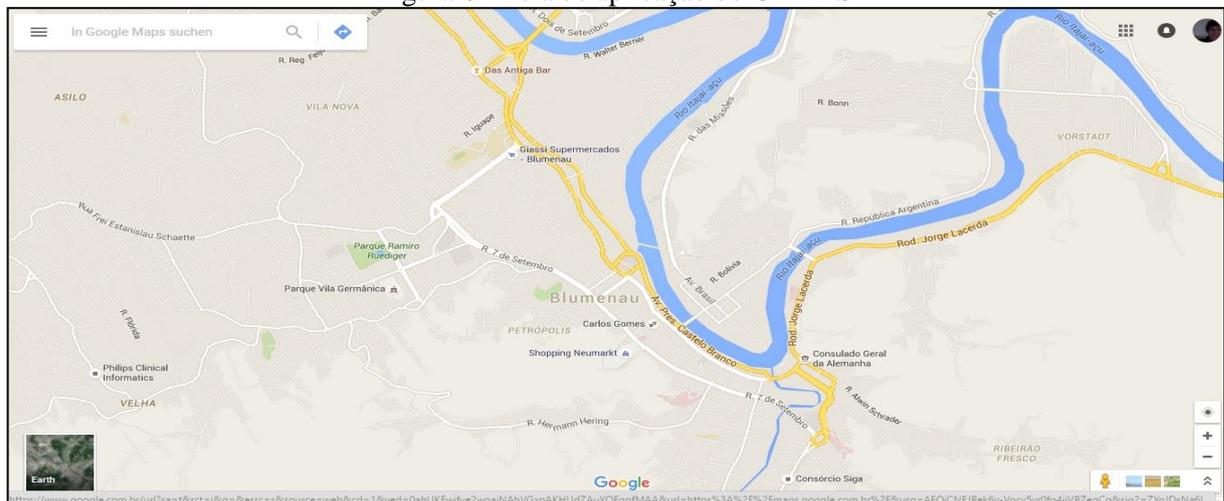
Pode se definir um SIG ou GIS em inglês (*Geographic Information System*), como um sistema de suporte e um conjunto de procedimentos para a obtenção, o armazenamento, a pesquisa, a análise, a representação, a visualização, a disponibilização e a publicação de dados geográficos. Estes dados podem ser representados por pontos, linhas, polígonos ou volumes. (COSME, 2012).

Uma das ferramentas SIG amplamente utilizada atualmente e de uso popular é o Google Maps (GMAPS). Esta ferramenta permite identificar pontos geograficamente e montar visualizações de agrupamento com mapa de calor. Nas próximas subseções, serão apresentadas informações sobre a ferramenta GMAPS e sobre mapas de calor.

2.3.1 Google Maps

O GMAPS é um serviço de pesquisa e visualização de mapas e imagens de satélites desenvolvido pela empresa estadunidense Google. O objetivo dessa ferramenta SIG é fornecer tecnologia geoespacial para todos, não apenas aos profissionais da área de tecnologia. A Figura 6 apresenta a tela de aplicação do GMAPS.

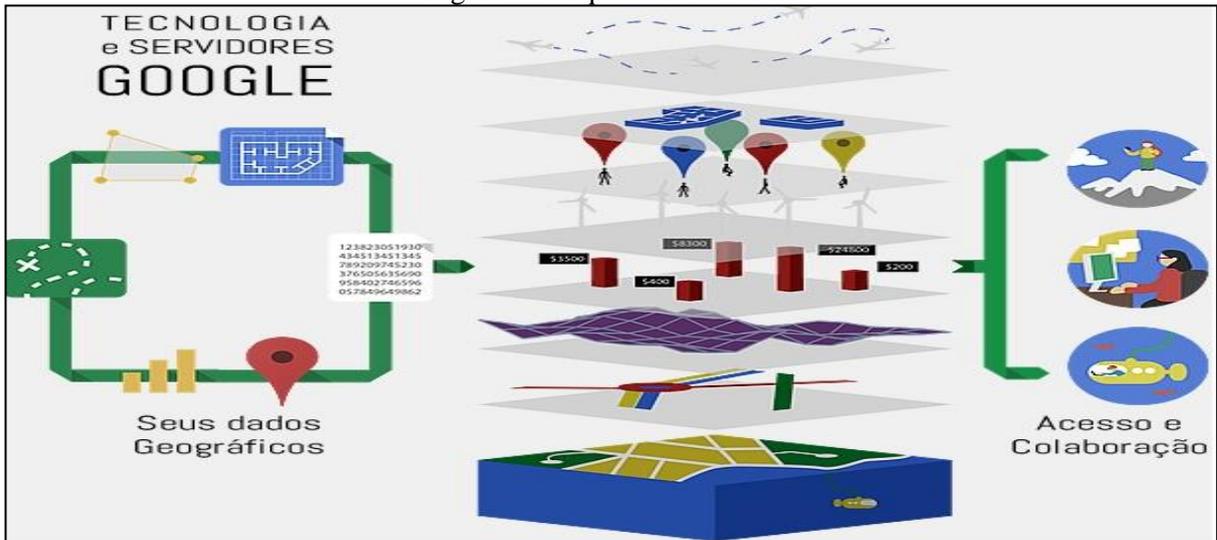
Figura 6 - Tela de aplicação do GMAPS



Fonte: Google Maps (2016).

O *Google Maps Engine* (GME) é a engrenagem desenvolvida para fornecer a mesma simplicidade e familiaridade do GMAPS e Google Apps. O GME permite fazer o *upload* de arquivos geográficos em formatos como *GeoTiffs* e *Shapefiles* e apresenta-los em camadas de mapas customizados ou criar grandes mosaicos de imagens. Na Figura 7 é demonstrado como é a arquitetura do GME (SCUSSEL, 2015).

Figura 7 - Arquitetura do GME



Fonte: Scussel (2015).

A Google oferece uma *Application Programming Interface* (API) em JavaScript para o desenvolvimento com a ferramenta do Google Maps, porém é necessário que o desenvolvedor tenha uma licença e uma chave para obter a API. Com essa API, é possível criar marcadores das localizações, criar rotas de caminhos usando diversos métodos de transporte, desenhar polígonos sobre o mapa, criar uma camada de mapa de calor e etc. (GOOGLE DEVELOPER, 2016).

Mapas de calores permitem que os visualizadores compreendam com mais facilidade a distribuição e a intensidade relativa de pontos em um mapa. Com grandes informações de dados, como de violência, é possível visualizar com mais facilidade o local com maior intensidade de violência. Na próxima seção, será abordada mais sobre os mapas de calor e também da biblioteca de mapa de calor do GMAPS.

2.3.2 Mapas de calor

Mapas de calor são uma das melhores ferramentas de visualização de dados de densidade de pontos. São utilizados para identificar facilmente aglomerados e encontrar onde existe uma elevada concentração de uma determinada atividade. Eles também são úteis para fazer análises de agrupamento ou análise de pontos de calor (*Hotspot Analysis*) (QGIS TUTORIALS AND TIPS, 2015).

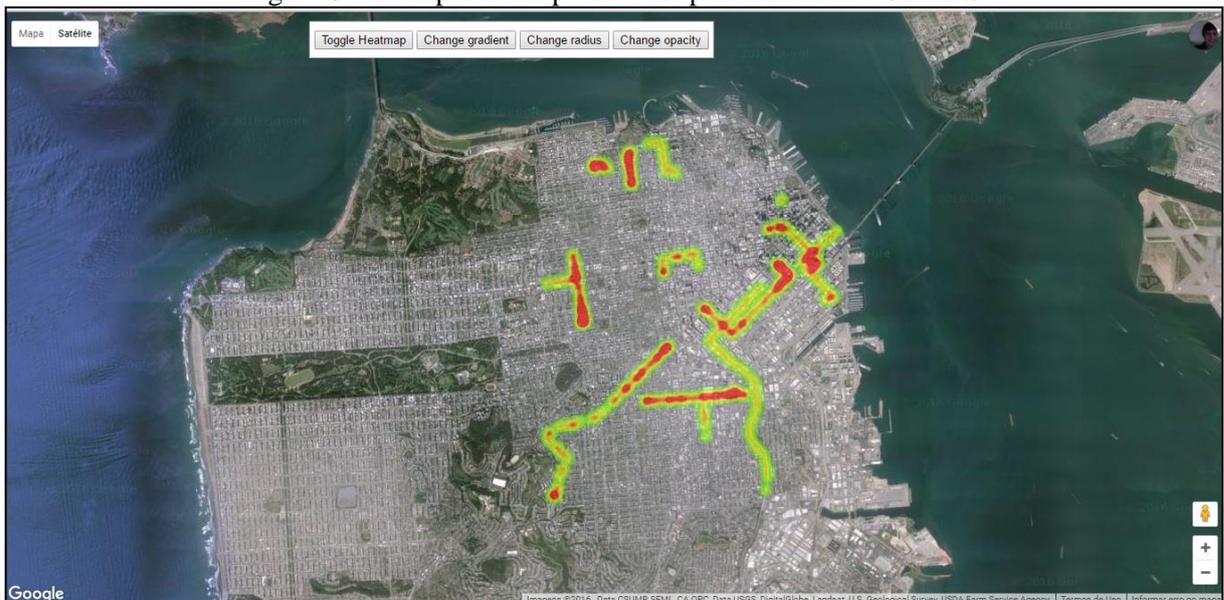
A ferramenta GMAPS, possui atualmente uma biblioteca para simular mapas de calor, chamada HeatMaps, no qual faz parte do pacote *google.maps.visualization*. Com ela é possível fazer uma sobreposição colorida sobre um mapa, diferenciando as áreas que tem mais intensidade dos dados em pontos geográficos. (GOOGLE DEVELOPER, 2016).

O HeatMaps possui configurações que permitem personalizar a renderização do mapa de calor. Abaixo algumas dessas configurações:

- a) *dissipating*: especifica se mapas de calor dissipam com o zoom;
- b) *gradient*: o gradiente de cor do mapa de calor, especificado como uma matriz de *Strings* de cores *Cascading Style Sheets* (CSS);
- c) *maxIntensity*: a intensidade máxima do mapa de calor. Por padrão, as cores do mapa de calor são dimensionadas dinamicamente de acordo com a maior concentração de pontos em qualquer *pixel* específico do mapa. Essa propriedade também permite especificar um máximo fixo;
- d) *radius*: o raio da influência de cada ponto de dados, em *pixels*;
- e) *opacity*: a opacidade do mapa de calor, expressa como um número entre 0 e 1.

A Figura 8 apresenta um exemplo de mapa de calor pela ferramenta do GMAPS.

Figura 8 - Exemplo de mapa de calor pela ferramenta GMAPS



Fonte: Google Developers (2016).

2.4 TRABALHOS CORRELATOS

Nesta seção são abordados três trabalhos correlatos. Na seção 2.4.1 é abordado o artigo científico de Bornhofen e Tenfen (2009). A seção 2.4.2 apresenta os artigos científicos de Furtado et al (2008) e Bordin et al (2013) e na seção 2.4.3 é apresentado um quadro comparativo dos trabalhos correlatos.

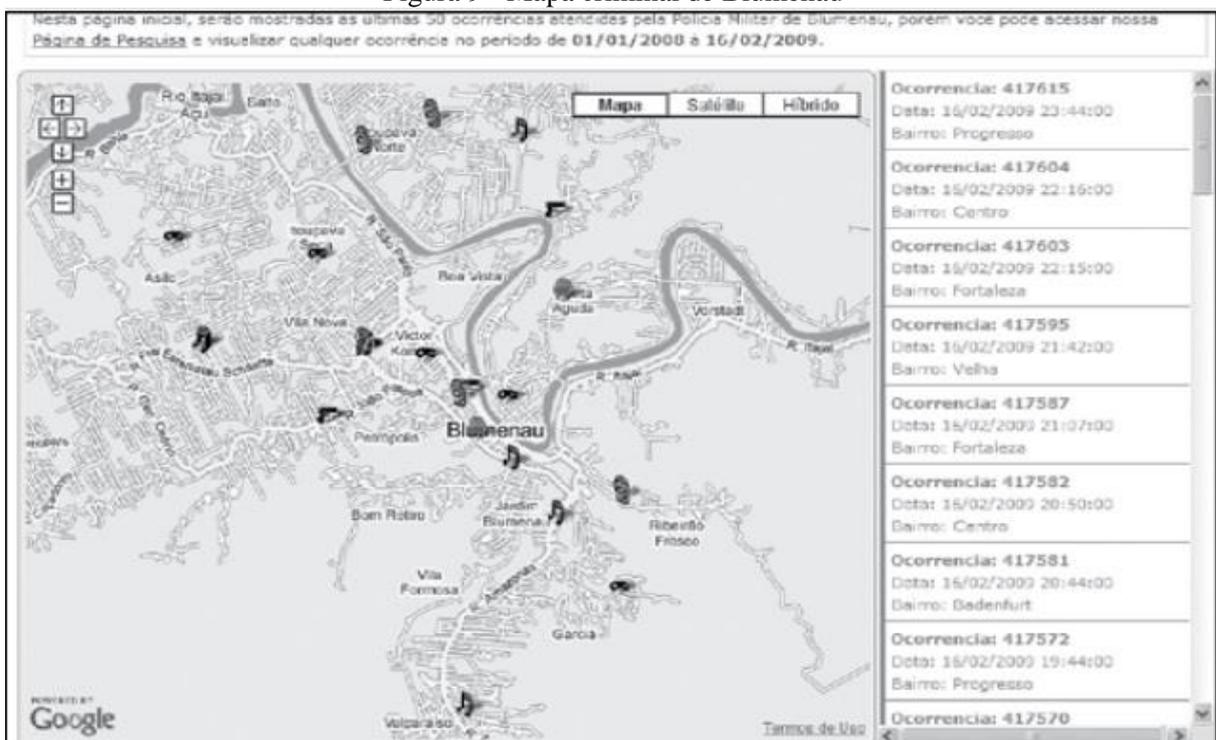
2.4.1 Mapeamento criminal por meio da plataforma Google Maps

O artigo científico de Bornhofen e Tenfen (2009) aborda o desenvolvimento e implementação de um sistema da PMB para a análise criminal utilizando a ferramenta GMAPS. O sistema tem como base a visualização de ícones sobrepostos no mapa que marcam graficamente as ocorrências policiais atendidas diariamente pela Polícia Militar em Blumenau. A ideia surgiu após os autores lerem reportagem na revista Exame sobre uma iniciativa de um cidadão nos Estados Unidos que montou um mapa com base nos dados que o departamento de polícia divulgava.

Para o desenvolvimento do trabalho, foi utilizado o banco de dados MySQL para o armazenamento das tabelas de ocorrências e naturezas das ocorrências. Também foi utilizado uma API do GMAPS para buscar a longitude e a latitude da ocorrência, bem como a apresentar geograficamente estas informações. O sistema foi desenvolvido na linguagem PHP e disponibilizado para visualização pública (10º BPM - Mapa de Ocorrências, 2008).

O sistema auxiliou no planejamento do policiamento da cidade, tanto no diário como nas operações que são implementadas. Também é destacado que a proposta é divulgar, de forma gratuita, os códigos da ferramenta para que qualquer organização policial militar possa elaborar seus mapas criminais sem custos (BORNHOFEN, 2009 apud LEMLE, 2009). Na Figura 9 é apresentado o mapa criminal de Blumenau no seu formato padrão de abertura.

Figura 9 - Mapa criminal de Blumenau



Fonte: Bornhofen e Tenfen (2009).

2.4.2 WikiCrimes – Um Sistema Colaborativo para Mapeamento Criminal

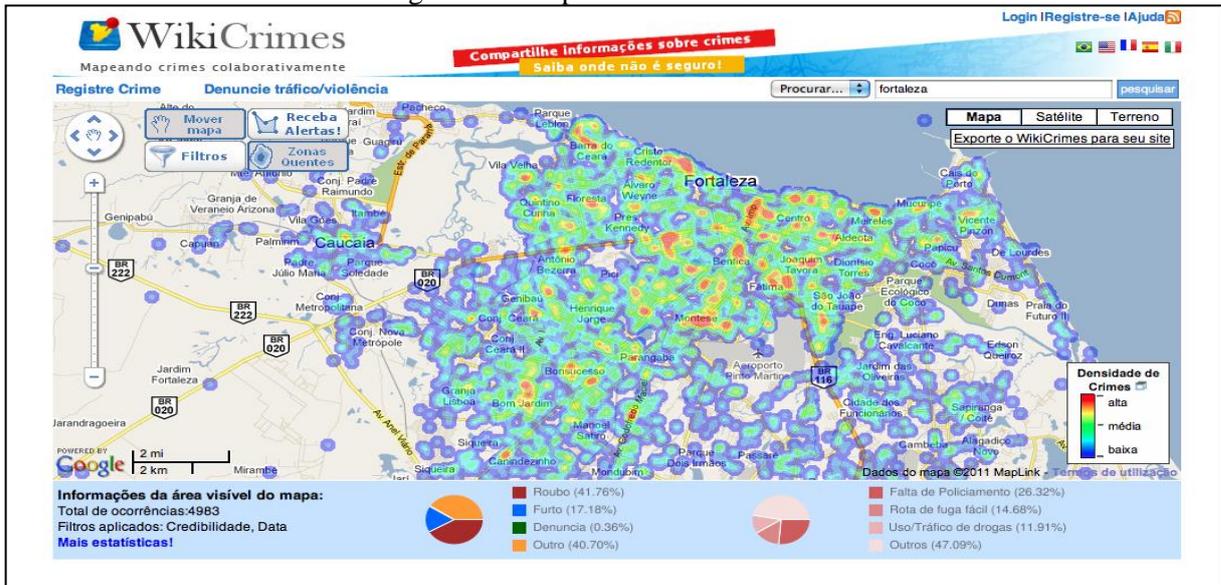
WikiCrimes é um software que permite o acesso e registro de ocorrências criminais no computador diretamente em um mapa digitalizado. O software segue a mesma lógica que a enciclopédia Wikipédia¹, deixando as informações sejam feitas de forma colaborativas por usuários e todos tendo o benefício de ter acesso às informações de crimes de mapa. Os exemplos mais chamativos desse software são alarmes e mapas de calor que indicam os lugares mais arriscados.

Para o desenvolvimento do trabalho, foram utilizados o banco de dados MySQL para o armazenamento de informações e a ferramenta Hibernate para a camada de integração com o banco de dados. Também foi utilizado os *frameworks* *JavaServer Faces* (JSF) e Springs para a fácil integração do cliente e servidor da aplicação. A *API* do GMAPS foi utilizada para a interação e manipulação usando classes do JavaScript. Com ela foi possível utilizar o *Geocoding*, onde foi possível realizar a busca por endereços, ruas, avenidas e etc.

O trabalho foi bem-sucedido e teve uma matéria internacional feita pela *British Broadcasting Corporation* (BBC) de Londres, onde entrevistou pessoas que foram a favor e contra o sistema. O departamento da Polícia Civil do Rio de Janeiro afirmou a BBC que tem preocupação de revelar dados que possam restringir seu trabalho e quer evitar criar uma sensação de insegurança à população. Para Vasco Furtado, ele acredita que a informação é exatamente ao contrário, onde cita que a insegurança é a falta de informação (CLEG, 2008). A Figura 10 demonstra a visualização de mapa de calor das áreas mais perigosas da cidade de Fortaleza no WikiCrimes.

¹ É uma enciclopédia livre web de amplo escopo que está a ser construída por milhares de colaboradores de todas as partes do mundo. Ela é baseada no conceito de Wiki Wiki, o que significa que qualquer internauta, pode editar o conteúdo de quase todos os artigos.

Figura 10 - Mapa de calor WikiCrimes



Fonte: Blog do Vasco (2011).

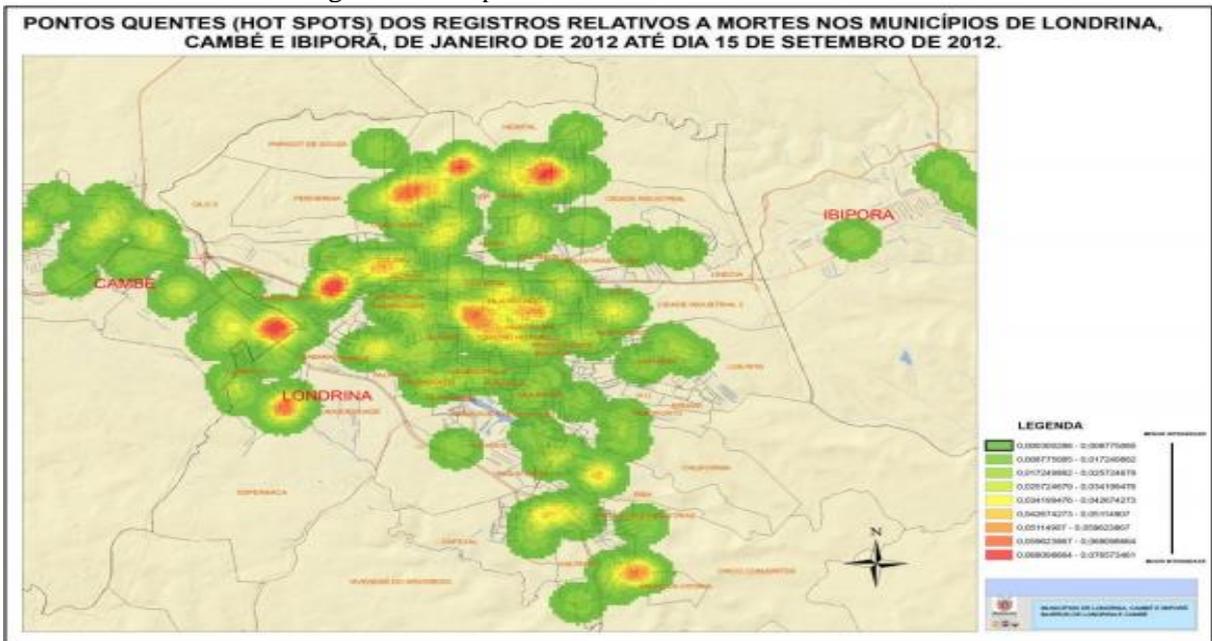
2.4.3 O uso do geoprocessamento na segurança pública do Estado do Paraná

O artigo de Bordin et al (2013) apresenta o projeto de desenvolvimento de um mapa do crime implantado pelo Estado do Paraná, através da Secretaria da Segurança Pública e por decisão do governador Roberto Requião no ano de 2003. O objetivo do projeto visava melhorar os resultados das duas polícias estaduais no desempenho das atividades de repressão e prevenção do crime e da violência.

O projeto começou com o desenvolvimento de um software chamado Boletim de Ocorrência Unificado (BOU), para armazenamento único de crimes e violências para as duas instituições policiais estaduais. Após o desenvolvimento do BOU, foi iniciado o desenvolvimento das informações cartográficas para a representação dos dados de crime. Para essa etapa foi realizado um convênio com a empresa Companhia de Energia Elétrica do Paraná (COPEL) e a Secretaria de Desenvolvimento Urbano (SEDU), para obter uma base de dados cartográfica e conseguir representar as informações de registros policiais.

Após realizada as duas etapas de desenvolvimento, foi possível avaliar os dados de registros policiais em um mapa georreferenciado por pontos, setores censitários, bairros ou por aglomerações de registros (densidade de *Kernell*) para a criação de *hot spots*. Na Figura 11 é apresentado os homicídios da cidade de Londrina de janeiro até setembro de 2012 com a densidade de *Kernell*.

Figura 11 - Mapa de homicídios cidade de Londrina



O projeto do desenvolvimento desse software trouxe à tona diversos problemas estruturais nas corporações policiais, em especial no tocante ao uso da técnica no tratamento espacial das informações sobre ocorrências policiais. Também sobre o uso da tecnologia da informação, uma vez que as informações não eram tratadas com a devida importância (BORDIN et al, 2013).

3 DESENVOLVIMENTO

As seções a seguir descrevem os requisitos, a especificação, a implementação e a operacionalidade do trabalho, abordando sucintamente as ferramentas e etapas utilizadas no desenvolvimento do projeto. Ao fim, são mostrados os resultados obtidos com este trabalho.

3.1 REQUISITOS

O SIG para monitoramento de ocorrências policiais deverá:

- a) disponibilizar uma interface web onde seja possível visualizar as áreas mineradas das ocorrências policiais através de um mapa de pontos e de calor (Requisitos Funcionais -RF);
- b) minerar as ocorrências policiais através de páginas de notícias de relatórios policiais (RF);
- c) coletar as ocorrências policiais através de relatórios anuais do Sistema Integrado de Segurança Pública (SISP)² (RF);
- d) buscar a coordenada geográfica das ocorrências policiais mineradas e coletadas (RF);
- e) classificar as ocorrências policiais encontradas como roubos, assaltos, homicídios, estupros, sequestros e etc. (RF);
- f) permitir ao usuário filtrar por período de data da ocorrência policial (RF);
- g) permitir ao usuário filtrar por classificação da ocorrência policial (RF);
- h) disponibilizar legendas de imagens para cada classificação de ocorrência policial no mapa de pontos (RF);
- i) utilizar o banco de dados MySQL para o armazenamento das ocorrências policiais e realizar a persistência dos dados através do framework Hibernate (Requisito Não Funcional - RNF);
- j) utilizar a API do GMAPS para o desenvolvimento de uma SIG (RNF);
- k) utilizar a linguagem de programação Java e utilizar a ferramenta para automação de compilação Maven (RNF);
- l) utilizar a linguagem de programação JavaScript (RNF);

² É um software desenvolvido pela Secretaria de Segurança Pública, junto com outros órgãos estaduais de segurança do estado de Santa Catarina, que armazena em um banco de dados todos os procedimentos policiais que são feitos e encaminhados ao Poder Judiciário.

- m) utilizar a biblioteca PTStemmer para lematizar as palavras para as classificações das ocorrências (RNF);
- n) utilizar a biblioteca JSOUP para extrair os dados do *site* Sala de Notícias (RNF);
- o) utilizar ambiente de programação Eclipse (RNF);
- p) estar acessível por navegadores mais populares, como por exemplo, o Chrome e o Firefox (RNF).

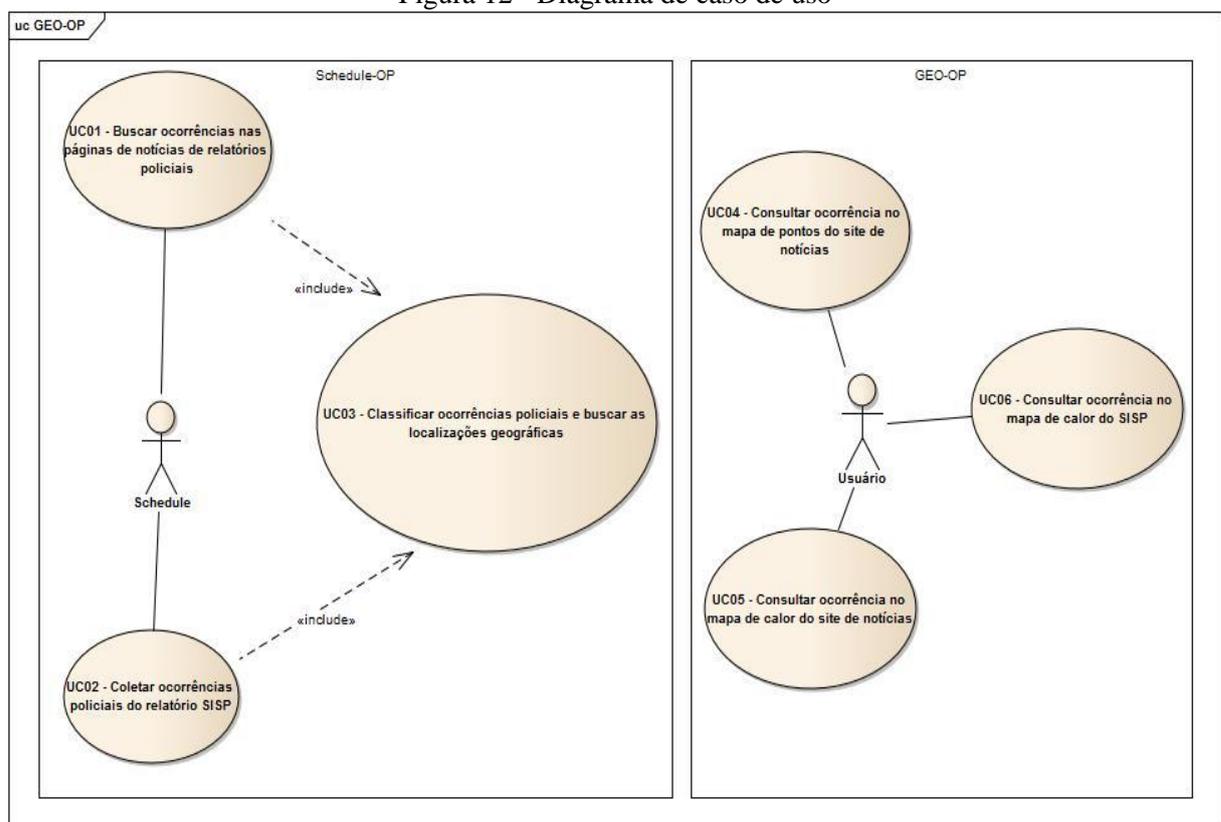
3.2 ESPECIFICAÇÃO

Para especificação do trabalho foi utilizada a ferramenta Enterprise Architect (EA) para a criação de diagramas de casos de uso, classes e pacotes.

3.2.1 Diagrama de casos de uso

Nessa seção são demonstradas as funções exercidas pelo sistema, sendo eles os atores Schedule e Usuário. A Figura 12 exibe o diagrama de caso de uso.

Figura 12 - Diagrama de caso de uso



Os casos de uso foram divididos de acordo com os sistemas que eles pertencem. O programa Schedule-OP está configurado como uma tarefa agendada no sistema operacional Windows para realizar as ações de coleta e classificações dos dados. Já o sistema do GEO-OP é

o sistema web da aplicação, onde apresenta mapas de pontos e de calor com os dados coletados e classificados pelo Schedule-OP.

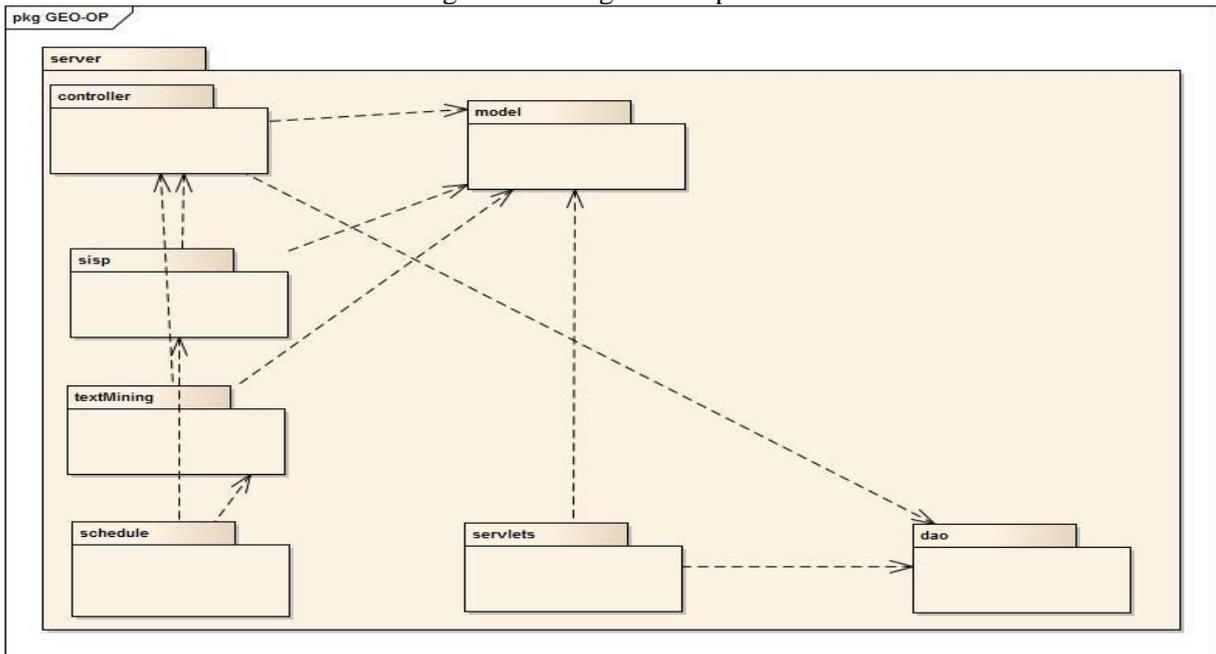
O primeiro caso de uso UC01 - Buscar ocorrências nas páginas de notícias de relatórios policiais é a ação feita pelo Schedule do sistema que buscar as ocorrências nas páginas de notícias de relatórios policiais no *site* Sala de Notícias (SALA DE NOTÍCIAS, 2012). O caso de uso UC02 - Coletar ocorrências relatório SISP, é a coleta das ocorrências dos relatórios policiais provenientes do sistema SISP. Após realizar as buscas das ocorrências pelos casos de UC01 e UC02, o Schedule executa o caso de uso UC03 - Classificar ocorrências policiais e buscar as localizações geográficas. Essa é a ação feita para classificar as ocorrências extraídas utilizando técnicas de MT, buscar a latitude e longitude do local da ocorrência e salvar as informações no Banco de Dados.

O caso de uso UC04 - Consultar ocorrência no mapa de pontos do site de notícias permite ao Usuário consultar no sistema GEO-OP ocorrências da cidade de Blumenau por pontos nos locais de ocorrências que foram coletados no *site* Sala de Notícias. O caso de uso UC05 - Consultar ocorrência no mapa de calor do site de notícias e UC06 - Consultar ocorrência no mapa de calor do SISP permite ao Usuário consultar as ocorrências da cidade de Blumenau por mapa de calor, com os dados do Sala de Notícias e SISP respectivamente, apresentando com a cor vermelha as áreas com maior número de ocorrência na região.

3.2.2 Diagrama de pacotes

Nessa seção são demonstrados a arquitetura de pacotes utilizados para o desenvolvimento do trabalho. A Figura 13 apresenta o diagrama de pacotes do sistema.

Figura 13 - Diagrama de pacotes

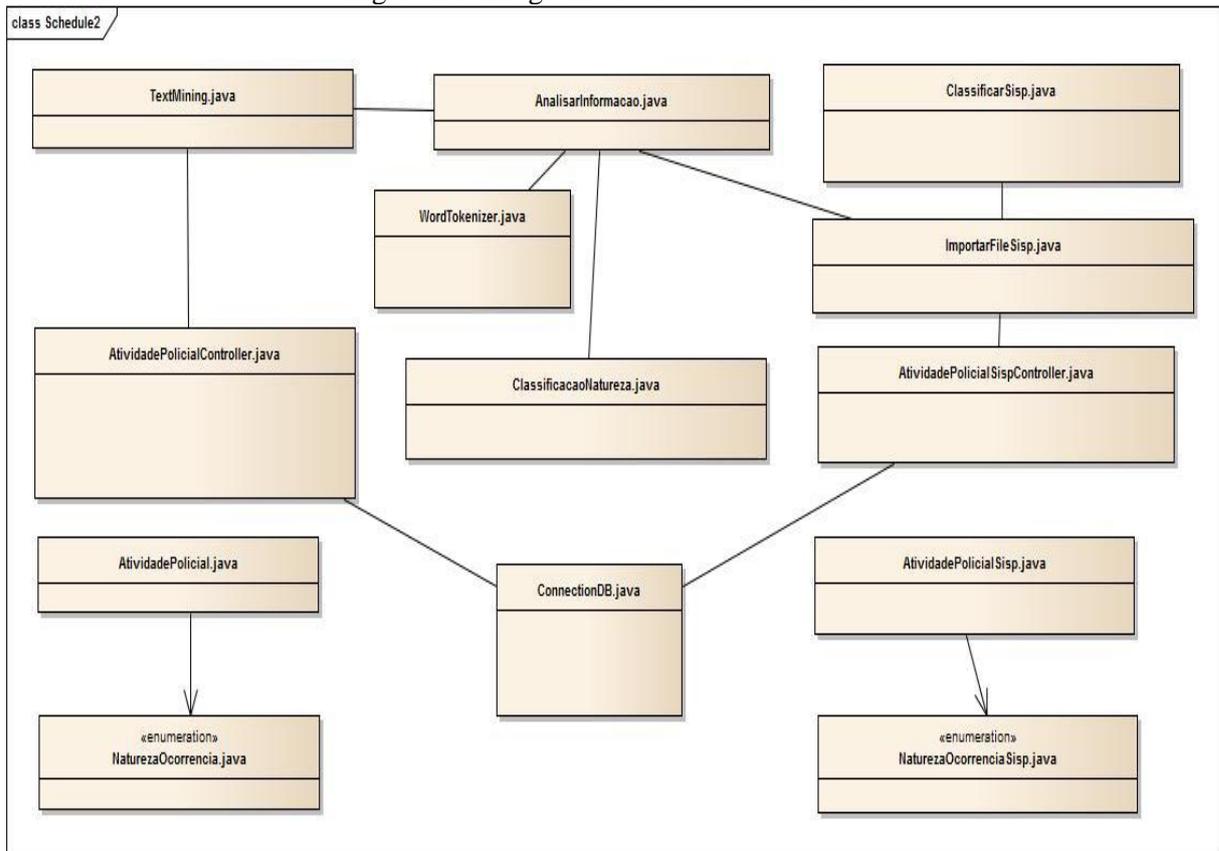


O pacote `schedule` é considerado o pacote “pai” dos demais, pois ele vai iniciar a aplicação do sistema para realizar a coleta e classificações das ocorrências. Os pacotes `sisp` e `textMining` são os responsáveis por fazer a intermediação do sistema, para as funções de extração de ocorrências das páginas de notícias e importação dos relatórios da polícia, busca da latitude e longitude das ocorrências e classificações das ocorrências, utilizando técnicas de MT. O pacote `controller` é utilizado para realizar as ações de controlador do sistema, onde envia comandos de inserção, alteração e exclusão dos registros de ocorrências extraídas no pacote `sisp` e `textMining`. O pacote `model` fica responsável pelo modelo de dados da aplicação, onde armazena classes com implementação com o Hibernate de criação de tabelas e domínios do sistema. O pacote `dao` é responsável pela conexão do banco de dados com a aplicação e por fim o pacote `servlets` é o pacote responsável por executar comandos recebidos pelo cliente da aplicação e retornar os dados da consulta das ocorrências.

3.2.3 Diagrama de classes

Nessa seção é abordado o diagrama de classes do sistema. A Figura 14 apresenta o diagrama de classes do Schedule do sistema sem a informação de métodos e atributos nas classes.

Figura 14 - Diagrama de classes do Schedule



Conforme já mencionado, o Schedule é responsável por iniciar a aplicação do sistema para buscas e classificações das ocorrências. Na classe `MiningText.java` são realizados os comandos de coleta de dados nas páginas de notícias, onde são coletadas as informações de ocorrências policiais. Após buscadas as ocorrências, é necessário realizar a separação de cada informação da ocorrência (data, fato e local) e para essa função a classe `AnalisarInformacao.java` é responsável. A classe `WordTokenizer.java` realiza a etapa de pré-processamento da MT e após realizada essa etapa a classe `AnalisaInformacao.java` chama a classe `ClassificacaoNatureza.java` para obter a classificação da ocorrência (roubo, furto, assalto e etc.).

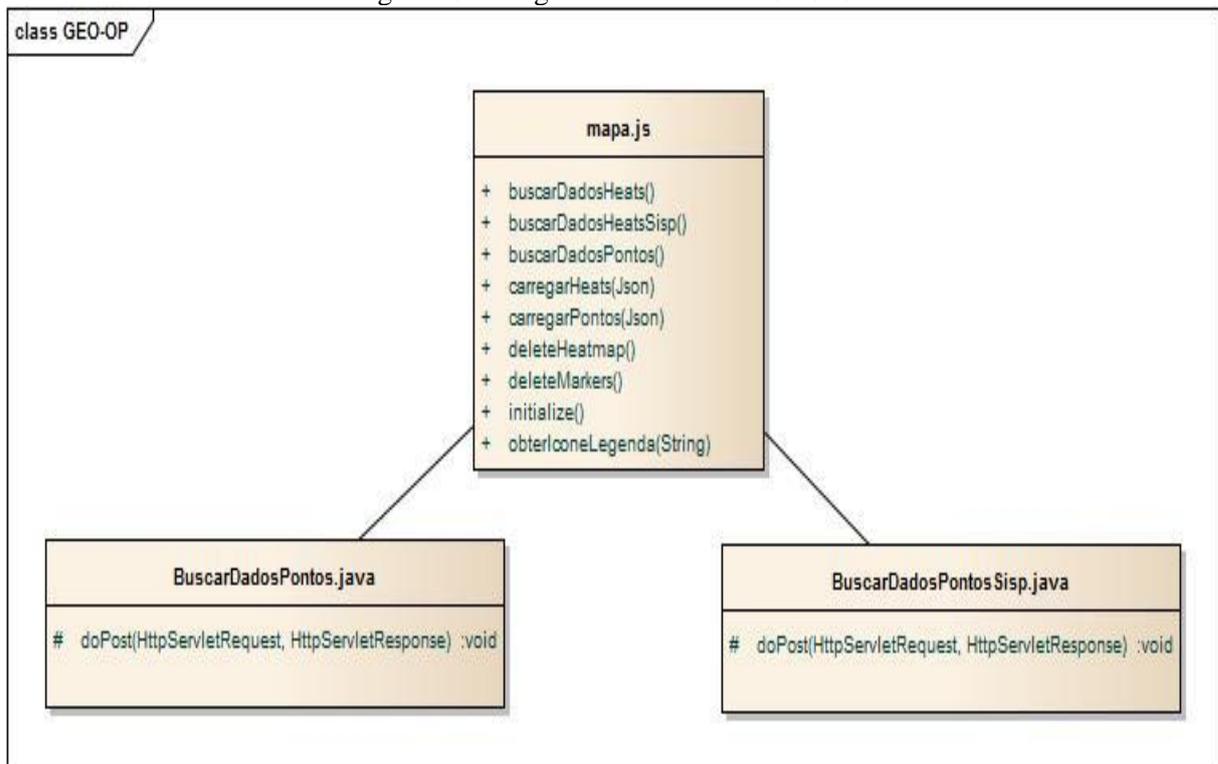
Para a importação dos relatórios da polícia o Schedule utiliza a classe `ImportarFileSisp.java`, onde recebe uma matriz de arquivos para realizar a importação dos relatórios. A classe `AnalisarInformacao.java` é aproveitada na importação dos relatórios para buscar a longitude e latitude da ocorrência. Para obter o tipo de classificação da ocorrência do relatório da polícia utiliza-se a classe `ClassificarSisp.java`.

As classes `AtividadePolicial.java` e `AtividadePolicialSisp.java` representam os objetos que serão persistidos e manipulados no banco de dados. A classe `AtividadePolicial.java` representa o modelo de dados para o armazenamento das

informações do site, onde é informado o fato, endereço, data, classificação, longitude e latitude da ocorrência. A classe `AtividadePolicialSisp.java` representa o modelo de dados para o armazenamento das informações dos relatórios da polícia. As classes `NaturezaOcorrencia.java` e `NaturezaOcorrenciaSisp.java` são *enums* onde armazenam domínios das classificações das ocorrências já definidas no sistema. As classes `AtividadePolicialController.java` e `AtividadePolicialSispController.java` são responsáveis por realizar a persistência e manipulação de dados. O Apêndice A apresenta o diagrama de classes do Schedule com os métodos e atributos.

A Figura 15 apresenta o diagrama de classe do Geo-OP. A classe `mapa.js` é responsável pela manipulação da API do GMAPS, dentro dessa classe ela requisita para as classes `BuscarDadosPontos.java` e `BuscarDadosPontosSisp.java` dados das ocorrências para a criação de mapas de calor e pontos via método *POST*. Essas classes também utilizam as classes `ConnectionDB.java`, `AtividadePolicial.java`, `AtividadePolicialSisp.java` do digrama de classes do Schedule para a manipulação de dados.

Figura 15 - Diagrama de classe do GEO-OP



3.3 IMPLEMENTAÇÃO

A seguir são mostradas as técnicas e ferramentas utilizadas e a operacionalidade da implementação.

3.3.1 Técnicas e ferramentas utilizadas

O ambiente de desenvolvimento utilizado para a implementação do sistema foi Eclipse Luna na versão 4.4.1. Essa ferramenta de desenvolvimento foi utilizada pois oferece uma série de facilidades para implementação do sistema, permitindo agilidade no desenvolvimento desse trabalho.

O desenvolvimento do trabalho foi dividido em 3 ciclos: O primeiro ciclo foi a coleta e análise de informações buscadas do *site* de notícias pelo sistema Schedule-OP. O segundo ciclo foi o desenvolvimento da importação e análise de dados dos relatórios da polícia, realizado pelo sistema Schedule-OP. O terceiro ciclo foi o desenvolvimento da aplicação do mapa *web*, que tem como objetivo representar os dados obtidos no primeiro e segundo ciclo de desenvolvimento, permitindo ao usuário visualizar e filtrar as ocorrências policiais. Nas seções seguintes são abordados os ciclos de desenvolvimento.

3.3.1.1 Primeiro ciclo de desenvolvimento

Neste ciclo de desenvolvimento, o sistema Schedule-OP coleta as ocorrências policiais através do *site* Sala de Notícias (SALA DE NOTÍCIAS, 2012), pois nele são publicados relatórios de atividades policiais diariamente da cidade de Blumenau. Na Figura 16 é demonstrada a página do *site* com *links* de notícias dos relatórios de atividades policiais.

Figura 16 - Relatórios de atividades policiais sala de notícias

The screenshot shows the website 'SALA DE NOTÍCIAS' with the following content:

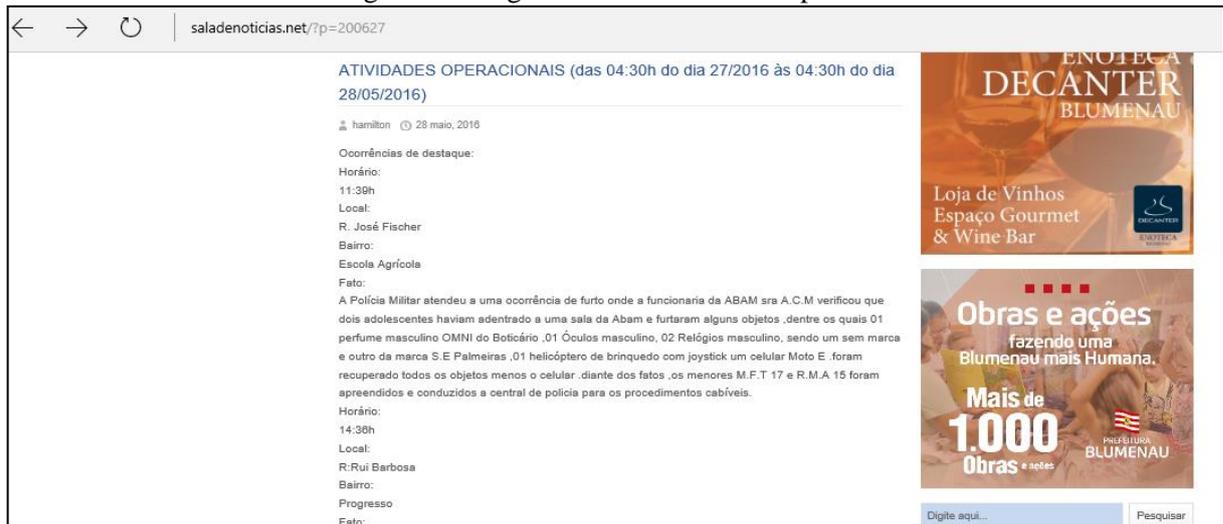
- Header: 'SALA DE NOTÍCIAS SOB A ÓTICA DE HAMILTON ANTONIO'
- Navigation: HOME, DESTAQUE, BLUMENAU, NACIONAL, SANTA CATARINA, INTERNACIONAL, ECONOMIA, SAÚDE, ENTRETENIMENTO, ESPORTE, CLASSIFICADOS
- Date: Quinta-Feira, 2 de Junho de 2016 - Ano 6
- Search Results for "REGIÃO POLICIAL MILITAR":
 - 7ª REGIÃO POLICIAL MILITAR CRE BLUMENAU RELATÓRIO DE ATIVIDADES OPERACIONAIS (das 04:30h do dia 27/05/2016 às 04:30h do dia 28/05/2016)
 - 7ª REGIÃO POLICIAL MILITAR CRE BLUMENAU RELATÓRIO DE ATIVIDADES OPERACIONAIS (das 04:30h do dia 26/2016 às 04:30h do dia 27/05/2016)
 - 7ª REGIÃO POLICIAL MILITAR CRE BLUMENAU RELATÓRIO DE ATIVIDADES OPERACIONAIS (das 04:30h do dia 22/05/2016 às 04:30h do dia 23/05/2016)
- Advertisements:
 - ENOTECA DECANTER BLUMENAU: Loja de Vinhos Espaço Gourmet & Wine Bar
 - Obras e ações: fazendo uma Blumenau mais Humana. Mais de 1000 Obras e ações.
- Search bar: Digite aqui... Pesquisar
- Footer: COLUNISTAS

Fonte: Sala de Notícias (2012).

A página disponibiliza diariamente relatórios policiais, permitindo ao leitor visualizar relatórios policiais antigos. Cada *link* de notícias do relatório de atividade policial possui várias informações de registros policiais, como o horário, local, data e a descrição do fato da

ocorrência. Conforme contato telefônico com o proprietário do *site*, essas informações são enviadas diariamente da Polícia Militar da cidade de Blumenau para que sejam publicadas no *site*, permitindo que estas tornam-se públicas para a população da região. Na Figura 17 é demonstrada a página do *site* com as ocorrências sobre um dos *links* de notícias dos relatórios de atividade policial.

Figura 17 - Página web de ocorrências policias



Fonte: Sala de Notícias (2012).

Para realizar a coleta de dados do *site*, foi utilizada a biblioteca JSOUP (JSOUPORG, 2009). Com essa biblioteca é possível fazer o *parse* de uma página HTML e achar conteúdos que estão dentro de atributos de HTML, como *href*, *div* e etc. O Quadro 1 apresenta um trecho de código onde é realizada o *parse* do HTML para o *link* de pesquisa dos relatórios de atividades policias.

Quadro 1 - Coleta de dados dos relatórios de atividades policias no *site* sala de notícias

```

01 Document document = Jsoup
02     .connect(
    "http://www.saladenoticias.net/?s=REGI%C3%83O+POLICIAL+MILITAR&submit=
        Pesquisar")
    .timeout(10 * 1000).get();
03 Elements searchResults = document.select(".post > h2 > a");
04 for (Element result : searchResults) {
05     String link = result.attr("href");
06     listaLinks.add(link);
07 }

```

Na linha 1 do Quadro 1 é criada a variável `document` onde irá conter todo o conteúdo da página do HTML, conforme URL informada na linha 2. Na linha 3 a variável `document` seleciona os elementos HTML que estão contidos em `.post > h2 > a` e busca todos os atributos `href` de *links* de ocorrências e salva em uma lista de *links*. O Quadro 1 apresenta a coleta de apenas as ocorrências de uma página de pesquisa, entretanto são mais de 90 páginas de conteúdo de pesquisa, com isso torna-se necessário realizar o *parse* dessas outras páginas

para coletar todas as ocorrências. O Quadro 2 apresenta a coleta das demais páginas de notícias.

Quadro 2 - Coleta de todos os *links* de ocorrências

```

01 for (int i = 2; i <= qtPages; i++) {
02     Document documentSub = Jsoup
03         .connect("http://www.saladenoticias.net/?s= REGI%C3%83O+
POLICIAL+MILITAR&submit=Pesquisar&paged="+ i).
04     timeout(10 * 1000).get();
05     Elements searchResultsSub = documentSub.select(".post > h2 > a");
06     for (Element resultSub : searchResultsSub) {
07         String link = resultSub.attr("href");
08         listaLinks.add(link);
09         System.out.println(link);
10     }
11     documentSub = null;
12     searchResultsSub.clear();
13 }

```

Na linha 1 do Quadro 2 é feito um laço de repetição com a variável `qtPages` que é uma variável com o valor 91 como padrão. Com o laço de repetição que se inicia do contador 2, são coletados 90 *links* armazenados em uma lista de *links*.

Após coletados todos os *links* de notícias do *site*, torna-se necessário entrar em cada *link* de notícia e coletar todas as ocorrências que estão associadas ao *link*. No Quadro 3 é demonstrada a coleta das ocorrências.

Quadro 3 - Coleta das ocorrências pelo *site*

```

01 Document document = Jsoup.connect(dsLink).timeout(10 * 1000).get();
02 Elements searchResults = document.select(".post-entry > p");
03 Elements searchDtOcorrencia = document.select(".heading-date");
04 String dsData = searchDtOcorrencia.text();
05 for (Element result : searchResults) {
06
07     if (result.text().trim().equalsIgnoreCase(" ")) {
08         continue;
09     }
10
11     if (result.text().equalsIgnoreCase("Ocorrências de destaque:")) {
12         continue;
13     }
14
15     if (result.text().contentEquals("UTILIDADE PÚBLICA")) {
16         break;
17     }
18
19     if ((result.text().length() > 17)
20         && (result.text().substring(0, 17)
21             .equalsIgnoreCase("UTILIDADE PÚBLICA"))) {
22         break;
23     }
25
26     extrairFatos(dsData, result.text());
27 }

```

No Quadro 3 é demonstrada a coleta pelas ocorrências dentro do *link* de notícias de atividades policiais. Quando é encontrada a informação de Utilidade pública no texto da

ocorrência, significa que informações de texto em diante não existe mais ocorrências e por isso é feito o comando *break* para que o algoritmo parar de processar a leitura do conteúdo do texto. Na linha 26 é realizada a extração dos dados das ocorrências, buscando o horário, local, bairro e o fato. No Quadro 4 é demonstrado um exemplo das ocorrências policiais do *site* que estão marcados entre os trechos em amarelo.

Quadro 4 - Exemplo de ocorrências policiais do *site*

<p>Ocorrências de destaque: Horário: 11:39h Local: R. José Fischer Bairro: Escola Agrícola Fato: ...</p> <p>Horário: 14:36h Local: R:Rui Barbosa Bairro: Progresso Fato:</p> <p>Horário: 16:12 h. Local: R: Araranguá ,678 Bairro: Garcia Fato: ...</p> <p>UTILIDADE PÚBLICA Solicitamos que os órgãos de imprensa deem ampla divulgação para a correta utilização do [...] Art. 265. Atentar contra a segurança ou o funcionamento de serviço de água, luz, força ou [...] Art. 340. Provocar a ação de autoridade, comunicando-lhe a ocorrência de crime ou de [...]</p>

Após realizada a coleta de dados, é necessário que seja feita a busca de localidade das ocorrências. Para esse processo foi utilizada a biblioteca GeoApiContext (GOOGLE DEVELOPER, 2016). Essa biblioteca foi desenvolvida pela Google e se comunica com um *WebService* para buscar a latitude e longitude, passando como parâmetro a informação do endereço desejado. No Quadro 5 é demonstrada a busca da latitude e longitude do local da ocorrência.

Quadro 5 - Busca da latitude e longitude da ocorrência

01	<code>GeocodingResult[] results = GeocodingApi.geocode(context,</code>
02	<code>dsLocal + ", Blumenau").await();</code>
03	
04	<code>for (GeocodingResult geocodingResult : results) {</code>
05	<code>setLatitude(geocodingResult.geometry.location.lat);</code>
06	<code>setLongitude(geocodingResult.geometry.location.lng);</code>
07	<code>break;</code>
08	<code>}</code>
09	<code>results = null;</code>

Na linha 1 do Quadro 5 é buscado o local da ocorrência concatenando a cidade de Blumenau, para que não apresente uma rua de outra cidade. Após isso, na linha 4 é aberto um comando de laço de repetição para buscar a latitude e longitude.

Para a classificação da ocorrência é necessário utilizar a técnica de MT chamada pré-processamento. Essa técnica retira os caracteres especiais, as *StopWords* (conforme listas de palavras de *StopWords* do Apêndice B) e o radical das palavras. Para a retirada dos radicais, foi utilizada a biblioteca PTStemmer (REUTEMANN, 2014). A biblioteca PTStemmer implementa três algoritmos de lematização de palavras da língua portuguesa, sendo eles: Orengo, Porter e Savoy. O Quadro 6 apresenta a técnica de pré-processamento sobre a descrição do fato da ocorrência.

Quadro 6 - Pré-processamento das ocorrências policias do *site* Sala de Notícias

```

01 String temp =
02     Normalizer.normalize(value, Normalizer.Form.NFKD).
           replaceAll("\\p{InCombiningDiacriticalMarks}+", "");
03 value = temp.replaceAll("[^\\p{ASCII}]", " ");
04 value = temp.replaceAll("\\p{Punct}]", " ");
05
06 stemmer = new OrengoStemmer();
07
08 String[] wordStrs = value.split("\\s");
09
10 ArrayList<Word> words = new ArrayList<Word>();
11 for (String word : wordStrs) {
12
13     if ((removeNumbersAndSingleLetters && word.length() > 1
14         && !word.matches("[0-9]+"))
15         || (!removeNumbersAndSingleLetters && word.length() > 0)) {
16
17         if (!stopWords.isStopWord(word)) {
18             word = word.replace(".", "");
19             word = word.replace(", ", "");
20             word = word.replace(")", "");
21             word = word.replace("(", "");
22             word = word.replace("\"", "");
23             word = word.replace("'", "");
24             words.add(new Word(stemmer.getWordStem(word)));
25         }
26     }
27 }
28 stemmer.clearIgnoreList();
29
30 return words;

```

Após feito o pré-processamento das palavras, é realizada as classificações das ocorrências, conforme a lista de classificações das páginas de notícias do Apêndice C. Para esse processo é feita uma análise textual com uma abordagem semântica, conforme explicado na seção 2.2.3, sobre o texto de cada ocorrência. Nessa análise é buscada primeira a ação da ocorrência, ou seja, procura-se um verbo, conforme a lista de verbos no Apêndice D. Após buscado o verbo é obtido qual foi o crime policial, conforme a lista de crimes policiais no Apêndice D. Para algumas ocorrências, é necessário ver qual foi o objeto policial para obter a classificação da ocorrência. No Quadro 7 é demonstrado um exemplo de uma descrição de ocorrência.

Quadro 7 - Exemplo de ocorrência policial do *site* Sala de Notícias

Exemplo 1: A Polícia Militar **atendeu** a uma ocorrência de **furto** onde a funcionaria da ABAM sra A.C.M verificou que dois adolescentes haviam adentrado a uma sala da Abam e furtaram alguns objetos, dentre os quais 01 perfume masculino OMNI do Boticário ,01 Óculos masculino, 02 Relógios masculino, sendo um sem marca e outro....

Exemplo 2: A Polícia Militar em rondas **abordou** o masculino J.N. de 16 anos, em busca pessoal foi **encontrado** em poder do mesmo 22 pedras de **Crack** e 05 buchas de **maconha**. Ante o exposto o adolescente foi apreendido e conduzido a Central de Polícia....

Conforme o primeiro exemplo do Quadro 7 é possível notar o verbo *atendeu*, prosseguido com o crime policial *furto*. Dessa forma o algoritmo classifica a ocorrência como **Furto**. Já no segundo exemplo, o verbo é *abordou* e o crime policial foi definido como *encontrado*. Porém com essas informações, não é possível classificar a ocorrência e consequentemente o algoritmo busca algum objeto no texto, conforme a lista de objetos de crimes policiais no Apêndice D. Nessa situação são encontradas as palavras *Crack* e *Maconha* e dessa forma o algoritmo classifica a ocorrência como **Porte de drogas**.

O Quadro 8 demonstra o passo-a-passo a classificação da ocorrência. Após receber as palavras pré-processadas (sem *StopWords*, radical da palavra e caracteres especiais), o algoritmo verifica, na linha 6, se a palavra é uma ação policial presente na lista de ações. Caso for, armazena a variável encontrada e busca o crime policial, na linha 14. Na linha 22 verifica-se se é necessário buscar o objeto policial e caso for, apura-se se alguma palavra está contida na lista de crimes policiais. Por fim, na linha 33, a aplicação chama outro algoritmo para classificar a ocorrência policial, conforme as palavras-chave encontradas nas linhas acima.

Quadro 8 - Classificação da ocorrência sala de notícias

```

01 Collection<Word> words = tokenizer.getWords(dsFato, stemmer);
02 int qtPassagens = 0;
03
04 for (Word word : words) {
05     if (qtPassagens == 0){
06         if (obterAcaoPolicial(word.toString())){
07             acao[qtPassagens] = word.toString();
08             qtPassagens++;
09         }
10         continue;
11     }
12
13     if (qtPassagens == 1){
14         if (obterCrimePolicial(word.toString())){
15             acao[qtPassagens] = word.toString();
16             qtPassagens++;
17         }
18         continue;
19     }
20
21     if ((qtPassagens == 2) &&
22         (obterSeVerificaObjeto(acao[1]))) {
23
24         acao[qtPassagens] = "nao-identifi";
25
26         if (obterObjetoPolicial(word.toString())){
27             acao[qtPassagens] = word.toString();
28             break;
29         }
30         continue;
31     }
32 }
33 return new ClassificacaoNatureza().getNaturezaOcorrencia(acao, dsFato);

```

3.3.1.2 Segundo ciclo de desenvolvimento

O segundo ciclo de desenvolvimento desse trabalho foi realizado a coleta de ocorrências dos relatórios de atividades policiais, enviados pela polícia e processados pelo sistema Schedule-OP. Os relatórios foram enviados a partir de uma requisição feita através de e-mail para o Gerente de Inteligência da Polícia Civil de Santa Catarina (PCSC) no modelo da Lei nº 12.527 (BRASIL, 2011). Foram recebidos 7 arquivos na extensão *.xlm* (*Excel*) com as informações de rua, bairro, data, hora e a tipificação da ocorrência. Entretanto, eles não disponibilizaram o texto da ocorrência, igual que existe no *site* de páginas de notícias.

Para realizar a importação desses arquivos ao sistema, foram convertidos os arquivos da extensão *.xlm* para a extensão *.csv*, onde cada informação de cada ocorrência é separada por ponto e vírgula. A Figura 18 demonstra o modelo em *.csv* do relatório de ocorrências da PCSC.

Figura 18 - Relatório de ocorrências policiares extensão .csv

Line	Record
22	2014;BLUMENAU;ÁGUA VERDE;FREI ESTANISLAU SCHAETTE;27/10/2014;10:00:00;Furto de telefone celular
23	2014;BLUMENAU;ÁGUA VERDE;GENERAL OSÁ"RIO;13/08/2014;10:10:00;Crimes eletrônicos (Lei n.º. 12.737/12)
24	2014;BLUMENAU;ÁGUA VERDE;GENERAL OSÁ"RIO;13/08/2014;10:10:00;Estelionato
25	2014;BLUMENAU;ÁGUA VERDE;ADHELE JENSEN;09/01/2014;12:00:00;Perturbação da tranquilidade
26	2014;BLUMENAU;ÁGUA VERDE;ÁGUA BRANCA;03/02/2014;14:00:00;Ameaça contra mulher (Violência doméstica)
27	2014;BLUMENAU;ÁGUA VERDE;ÁGUA BRANCA;03/02/2014;14:00:00;Lesão corporal dolosa mulher (Violência doméstica)
28	2014;BLUMENAU;ÁGUA VERDE;ÁGUA BRANCA;07/02/2014;11:45:00;Danos
29	2014;BLUMENAU;ÁGUA VERDE;ÁGUA BRANCA;15/04/2014;11:10:00;Cumprimento de Mandado de Prisão
30	2014;BLUMENAU;ÁGUA VERDE;ÁGUA BRANCA;21/03/2014;;Fato atípico
31	2014;BLUMENAU;ÁGUA VERDE;ÁGUA VERDE;21/02/2014;16:00:00;Ameaça contra homem
32	2014;BLUMENAU;ÁGUA VERDE;ÁGUA VERDE;25/03/2014;21:00:00;Jogos de azar
33	2014;BLUMENAU;ÁGUA VERDE;ALEIXO SCHNEIDER;21/02/2013;19:00:00;Suicídio de mulher
34	2014;BLUMENAU;ÁGUA VERDE;ALEXANDRE DE SOUZA;09/03/2014;18:00:00;Danos (Violência doméstica)
35	2014;BLUMENAU;ÁGUA VERDE;ALEXANDRE DE SOUZA;09/03/2014;18:00:00;Injúria (Violência doméstica)
36	2014;BLUMENAU;ÁGUA VERDE;ALEXANDRE DE SOUZA;09/03/2014;18:00:00;Lesão corporal dolosa mulher (Violência doméstica)
37	2014;BLUMENAU;ÁGUA VERDE;ALEXANDRE DE SOUZA;23/03/2014;16:00:00;Apresentação de pessoa na Delegacia por Agente Público
38	2014;BLUMENAU;ÁGUA VERDE;ALEXANDRE DE SOUZA;23/03/2014;16:00:00;Cumprimento de Mandado de Prisão
39	2014;BLUMENAU;ÁGUA VERDE;ALFERES JULIO SAMETZKI;15/08/2014;;Ameaça contra mulher
40	2014;BLUMENAU;ÁGUA VERDE;ALFERES JULIO SAMETZKI;15/08/2014;;Injúria
41	2014;BLUMENAU;ÁGUA VERDE;ALFERES JULIO SAMETZKI;18/03/2014;09:00:00;Ameaça contra homem
42	2014;BLUMENAU;ÁGUA VERDE;ALFERES JULIO SAMETZKI;18/03/2014;09:00:00;Ameaça contra mulher
43	2014;BLUMENAU;ÁGUA VERDE;ALFERES VON SECKENDORF;20/01/2014;;Ameaça contra mulher (Violência doméstica)
44	2014;BLUMENAU;ÁGUA VERDE;ALFERES VON SECKENDORF;21/08/2014;02:00:00;Furto de automóvel/camioneta
45	2014;BLUMENAU;ÁGUA VERDE;ALFONSO SOUZA E SILVA;20/02/2014;16:55:00;Ameaça contra homem
46	2014;BLUMENAU;ÁGUA VERDE;ALMIRANTE TAMANDARÉ;20/03/2014;09:40:00;Furto em residência
47	2014;BLUMENAU;ÁGUA VERDE;ALWIN BECKMANN;;Ameaça contra mulher (Violência doméstica)
48	2014;BLUMENAU;ÁGUA VERDE;AMELIA VENSKE;18/07/2014;18:00:00;Difamação
49	2014;BLUMENAU;ÁGUA VERDE;AMELIA VENSKE;18/07/2014;18:00:00;Perturbação do trabalho ou sossego alheio
50	2014;BLUMENAU;ÁGUA VERDE;ANASTACIO RAYMUNDO DA SILVA;;Fato atípico
51	2014;BLUMENAU;ÁGUA VERDE;ANASTACIO RAYMUNDO DA SILVA;08/03/2014;01:00:00;Furto de documento
52	2014;BLUMENAU;ÁGUA VERDE;ANGELINA;07/04/2014;17:00:00;Lesão corporal dolosa mulher (Violência doméstica)
53	2014;BLUMENAU;ÁGUA VERDE;APOLONIA VON BUETNER;09/04/2014;20:00:00;Jogos de azar
54	2014;BLUMENAU;ÁGUA VERDE;ARNOLDO PRIM;02/03/2014;21:15:00;Furto de automóvel/camioneta
55	2014;BLUMENAU;ÁGUA VERDE;ARNO DO BUEDEFFER;25/02/2014;10:40:00;Estelionato

Fonte: Relatório de ocorrências fornecidos por e-mail pela PCSC.

Como já mencionado, a descrição da ocorrência não foi enviada pela PCSC, devido a essa falta de informação é realizada a classificação da ocorrência conforme a última coluna da Figura 18. O código fonte do Quadro 9 demonstra a forma de classificação das ocorrências dos relatórios policiais. Na linha 2 é utilizado o mesmo algoritmo apresentado no Quadro 6 para realizar a etapa de pré-processamento da MT. Após realizada a análise nas palavras, é buscado a palavra do crime policial da lista de crimes, conforme a lista de crimes policiais no Apêndice D do Quadro 21.

Quadro 9 - Classificação das ocorrências de relatórios policiais

```

01 String acao = "nao-identifi";
02 Collection<Word> words = tokenizer.getWords(dsFato, stemmer);
03
04 for (Word word : words) {
05
06     if (obterCrimePolicial(word.toString())) {
07         acao = word.toString();
08         break;
09     }
10     continue;
11 }
12 crime = getNaturezaSisp(acao);

```

3.3.1.3 Terceiro ciclo de desenvolvimento

Nesse terceiro e último ciclo do desenvolvimento, será apresentado a forma com que foi construída a tela de mapas *web* do sistema GEO-OP. A tela *web* foi construído utilizando

HTML5 e a ferramenta Bootstrap (Bootstrap, 2016) para o *desing* da tela. Para o desenvolvimento dos mapas de pontos e de calor foi utilizada a API do GMAPS (GOOGLE DEVELOPER, 2016) com a linguagem JavaScript.

As requisições para buscar os dados das ocorrências foram consumidas via AJAX por uma *Servlet* no lado servidor da aplicação. O Quadro 10 apresenta a requisição do AJAX para a *Servlet* do servidor via método *POST*.

Quadro 10 - Requisição via Servlet do Servidor

```

01 $.ajax({
02     url: " saegopBuscarDadosSisp",
03     type: 'POST',
04     data: {'dt-inicio' : dtInicio.value, 'dt-fim' : dtFim.value,
'id-Crime' : document.getElementById("idCrime").value},
05     dataType: "json",
06     success: function (data) {
07         carregarHeats(data);
08         return false;
09     },
10     error:function(data,status,er) {
11         alert("error: "+data+" status: "+status+" er:"+er);
12     }
13 });

```

Na linha 2 até 5 é configurada a chamada *Ajax* para o servidor com as informações do tipo de requisição, o endereço do serviço e tipo do conteúdo que será enviado. Os dados da requisição são retornados no formato *JavaScript Object Notation* (JSON) onde, posteriormente, carregará no mapa de calor ou de pontos as informações das ocorrências policiais.

Para a criação do mapa de calor é necessário utilizar a biblioteca *HeatmapLayer* do GMAPS. No Quadro 11 é demonstrada a criação do mapa de calor no sistema.

Quadro 11 - Criação de mapa de calor

```

01 var pointsPosition = [];
02
03 $.each(pontosJson, function(index, ponto) {
04     position = new google.maps.LatLng(ponto.latitude,
ponto.longitude);
05     pointsPosition.push(position);
06 });
07
08 var heatmap = new google.maps.visualization.HeatmapLayer({
09     maxIntensity: 50,
10     data: pointsPosition
11 });
12
13 heatmap.set('radius', heatmap.get('radius') ? null : 20);
14 heatmap.set('gradient', heatmap.get('gradient') ? null : gradient);

```

O trecho entre as linhas 3 até 6 apresenta um laço de repetição com todos os dados de ocorrências que estão na variável *pontosJson* que contém informações do tipo JSON e são

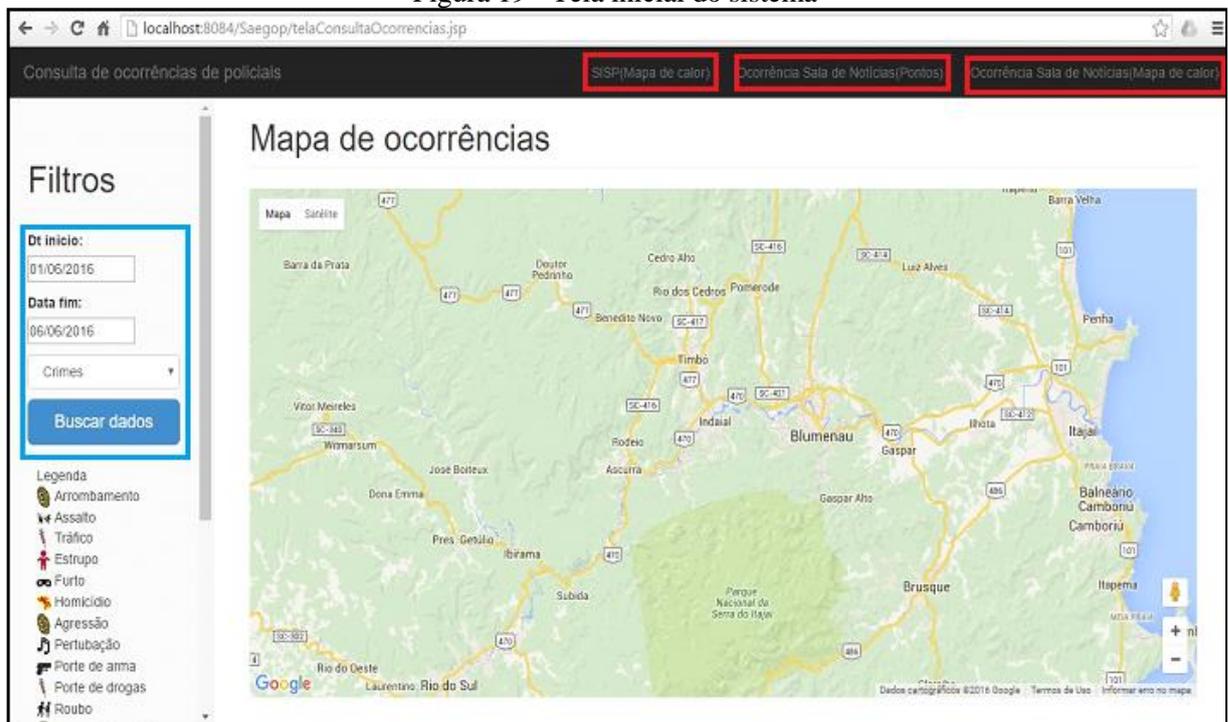
associados a uma matriz de pontos de localização chamada `pointsPosition`. O trecho entre as linhas 8 até 11 é criado o mapa de calor utilizando a biblioteca `HeatmapLayer` e passando os dados da matriz de pontos `pointsPosition`. Nas linhas 13 e 14 são definidos o raio da influência de cada ponto de dados e o gradiente de cor do mapa de calor.

3.3.2 Operacionalidade da implementação

Nesta seção é apresentada a operacionalidade da implementação, com as suas funcionalidades no sistema GEO-OP. São demonstrados os tipos de consulta das ocorrências por mapa de calor e de pontos.

A Figura 19 apresenta a tela inicial do GEO-OP, com as informações de filtros (circulado com a cor azul) de `Dt início`, `Dt fim` e `Crimes` e também o botão `Buscar dados`, para a busca de dados e montagem do mapa das ocorrências. Na parte superior da tela (circulado com a vermelha) é possível visualizar os tipos de mapas, como o mapa de calor do `site` notícias, o mapa de calor do SISP (dados dos relatórios de ocorrências policiais enviadas pela PCSC) e o mapa de pontos do `site` de notícias, que é o mapa principal ao abrir a tela do `site`. O mapa de pontos com os dados do SISP não foi desenvolvido, pois como mencionado, a PCSC não enviou as informações da ocorrência e dessa forma iria ficar diferente do mapa de pontos do `site` de notícias.

Figura 19 - Tela inicial do sistema



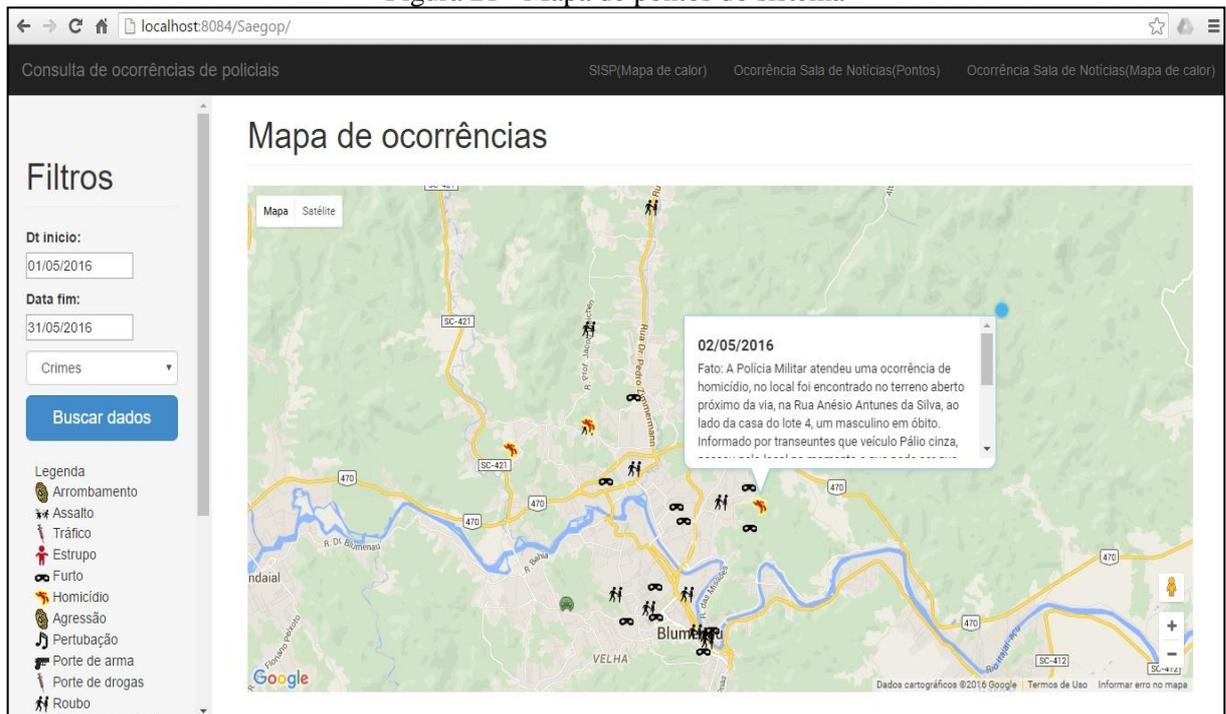
A Figura 20 apresenta as legendas de crimes do mapa de pontos do *site* de notícias na parte esquerda da tela. Cada legenda possui uma ilustração de cada classificação de crime do *site* de notícias.

Figura 20 - Legendas do mapa de pontos do *site* de notícias



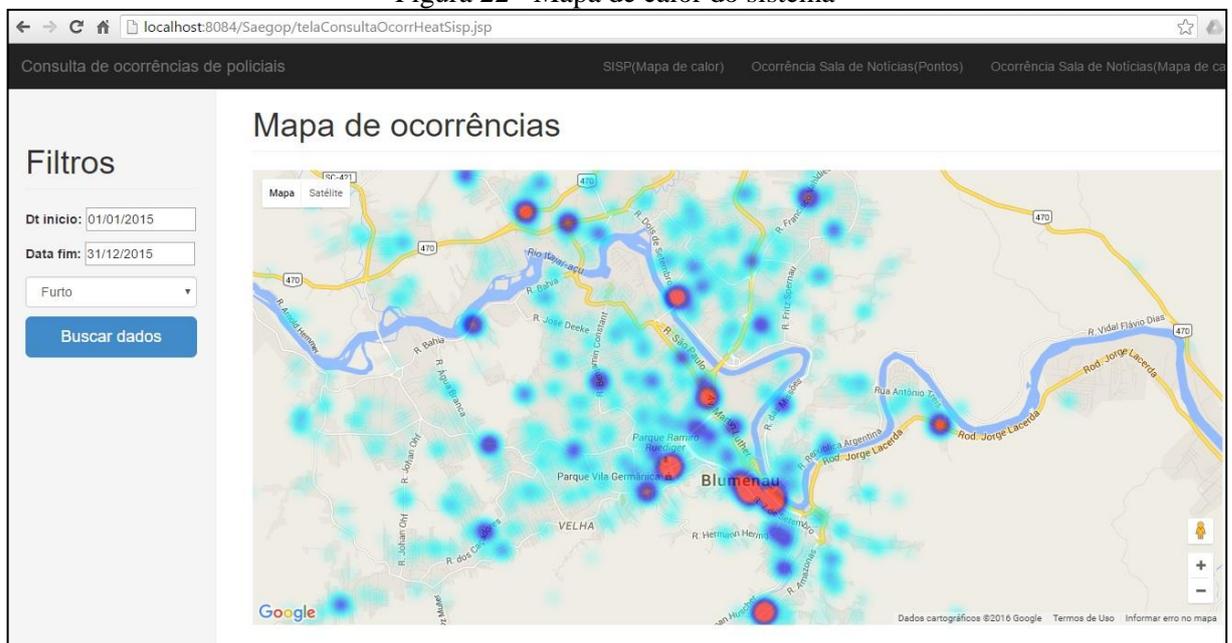
A Figura 21 apresenta a busca de ocorrências de 01/05/2016 a 31/05/2016, onde são apresentadas as ocorrências no mapa de pontos do *site* de notícias. Cada ponto de ocorrência é representado por uma legenda diferente, conforme o seu crime. Ao clicar no ponto, são apresentadas as informações da data e o fato da ocorrência.

Figura 21 - Mapa de pontos do sistema



A Figura 22 apresenta o mapa de calor do relatório policial, enviado pela PCSC, do ano de 2015, apenas dos crimes de furtos.

Figura 22 - Mapa de calor do sistema



3.4 RESULTADOS E DISCUSSÕES

Este trabalho permite observar a violência na cidade de Blumenau utilizando um SIG e técnicas de MT nas coletas e classificações de dados. Com base nessas informações, é

possível identificar a viabilidade de utilizar ambas essas técnicas afim de observar as áreas mais violentas da cidade.

Na seção 3.4.1 são apresentados os resultados da coleta e classificação de dados nas páginas de notícias, na seção 3.4.2 são apresentados os resultados da coleta e classificação de dados dos relatórios policiais enviados pela PCSC. Na seção 3.4.3 são apresentadas as regiões da cidade com maiores índices de violência e por fim na seção 3.4.4 é feito uma comparação com os trabalhos correlatos e o sistema desenvolvido.

3.4.1 Resultados das coletas e classificações dos dados nas páginas de notícias

Foram coletados cerca de 4234 registros e classificadas 90% das ocorrências nas páginas de notícias usando o algoritmo semântico criado. Também foram encontrados cerca de 95% das localizações das ocorrências, ou seja, foi encontrado a latitude e longitude. A Tabela 1 apresenta todos os tipos de ocorrências classificadas junto com a respectiva quantidade de registros, destacando que não foi possível identificar se o algoritmo classificou corretamente as ocorrências, visto que os dados do *site* não possuem uma classificação prévia.

Tabela 1 – Classificações das ocorrências realizadas pelo algoritmo no *site*

Ocorrência	Quantidade
Furto	2042
Roubo	743
Arrombamento	238
Não identificado	226
Não possui fato	213
Porte de Drogas	184
Mandato de prisão	107
Lesão Corporal/Agressão	94
Encontrado veículo roubado	84
Assalto	69
Porte de Arma	48
Homicídio	41
Disparo de arma de fogo	41
Embriaguez	26
Tráfico	26
Perturbação	19
Prisão de foragido	15
Pagamento com moeda falsa	6
Encontrado cadáver humano	5
Estelionato	5
Estupro	2
	4234

Não foram encontradas a localização (latitude e longitude) de cerca de 235 registros de ocorrências. Esse problema representou cerca de 5% da base, considerando que muitas das ocorrências nessas ruas estão descritas no *site* há mais de 2 anos. As ruas com maiores

registros de localização não encontradas possuem as seguintes descrições: R Eng Udo, Via Expressaressa Paul, R Jacob Bornhofen, R Jos Seibt, R Augusto Grohman, Eng Paul Werner, R Osni Jensen. Ao todo elas somam em 33 registros de ocorrências. Um dos principais motivos para a biblioteca do Google não encontrar essas ruas deve-se a descrição errada da rua ou abreviações, como **Expressaressa** e **R Eng Udo**.

Houve também problemas ao coletar as descrições dos fatos das ocorrências, onde cerca de 213 registros de fatos de ocorrências não foram encontrados. Esse problema foi gerado devido a erros de HTML da página do *site*. Porém, esse problema representou cerca de 5% da base e com isso não foi possível classificar essas ocorrências. Também não foram classificadas cerca de 226 ocorrências devido a problemas do algoritmo de classificação semântico e o texto da ocorrência. Um dos exemplos está no Quadro 12, onde é possível visualizar a palavra *recuperou* é o primeiro verbo apresentada no texto do fato da ocorrência e logo depois a palavra *veículo*. Entretanto a palavra *possuía* que é considerada como crime policial é colocada depois, onde posteriormente o algoritmo iria buscar o objeto policial, que nesse caso seria a palavra *veículo*. Com essas situações, o algoritmo não conseguiu classificar essa ocorrência como: Encontrado veículo roubado.

Quadro 12 - Ocorrências não classificadas do *site*

Fato: A Polícia Militar **recuperou** o **veículo** GM/CORSA WIND, de cor branca, ano 1997, placa LBR5330 de Navegantes, que **possuía** registro de furto naquela cidade, removido a Delegacia de Polícia para as devidas providências.

Algumas ocorrências também foram classificadas erradas, devido a apresentação do crime policial ter sido apresentado primeiro que o crime policial original da palavra. O Quadro 13 apresenta um exemplo de classificação errada do algoritmo.

Quadro 13 - Classificação da ocorrência errada do *site*

Fato: A Polícia Militar atendeu uma ocorrência de **furto** em estabelecimento comercial, onde foi **arrombada** a porta da imobiliária XXX, e subtraído do local uma CPU da marca Space BR de cor preta e um monitor da marca OAC na cor preta, após a informação recebida do arrombamento via 190...

A palavra *furto* é um crime policial, porém logo vem a palavra *arrombada*, no qual o crime se caracterizasse como um arrombamento. Nessa situação o algoritmo acabou classificando a ocorrência como furto, pois a primeira palavra de crime policial é *furto* e não *arrombada*.

3.4.2 Resultados das coletas e classificações dos dados dos relatórios policiais

Foram coletados cerca de 107403 registros dos relatórios policiais enviados pela PCSC e classificados 97% ocorrências. Também foram encontrados cerca de 88% das localizações das ocorrências, ou seja, foi encontrado a latitude e longitude. A Tabela 2 apresenta o resultado final de todas as ocorrências classificadas pelo algoritmo. É importante observar que a Tabela 2 possui uma maior quantidade de tipos de ocorrências que a Tabela 1 do *site* de notícias, pois os registros coletados pelos relatórios policiais possuíam ocorrências previamente classificadas e também uma quantidade maior de dados.

Tabela 2 - Classificações das ocorrências realizadas pelo algoritmo nos relatórios da polícia

Ocorrência	Quantidade
Furto	21789
Ameaça	19774
Lesão Corporal/Agressão	13434
Fato atípico	7098
Injúria	6738
Dano	6472
Estelionato	4376
Não identificado	4149
Roubo	2963
Perturbação	2845
Posse/Tráfico de drogas	2143
Acidente de trânsito	2062
Embriaguez	1763
Difamação	1573
Calúnia	1425
Desacato/Desobediência pela autoridade	1381
Homicídios e tentativas/Suicídios/Mortes adversas	1206
Fuga	1042
Mandato de prisão	831
Arrombamento	769
Estupro e tentativa	724
Desaparecimento de pessoa	677
Abandono de incapaz	602
Crime sexual	357
Reaparecimento de pessoa	355
Crimes eletrônicos	335
Disparo / Porte de arma de fogo	293
Incêndio	227
	107403

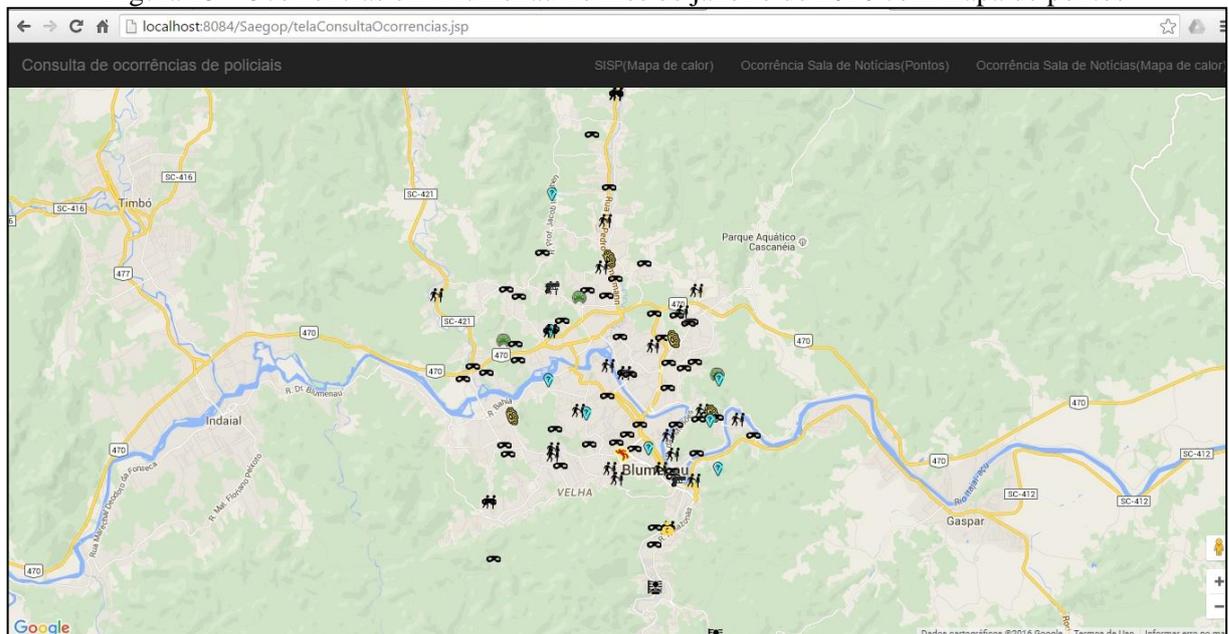
Não foram encontradas a localização (latitude e longitude) de cerca de 13712 dados. Cerca de 4358 ocorrências foram enviadas a descrição da rua, como: “NÃO INFORMADO”, “DESCONHECIDO” ou nula. Algumas ruas também foram enviadas com a informação errada no nome, como por exemplo: RUA BAIRRO GLORIA, PASTOR OSWALD HESSE.

Para as classificações das ocorrências, cerca de 4149 não foram classificadas, devido à falta de informação em algumas ocorrências, como por exemplo: “Apresentação de pessoa na Delegacia por Agente Público” e “Exercício arbitrário das próprias razões”.

3.4.3 Regiões com maiores índices de violências

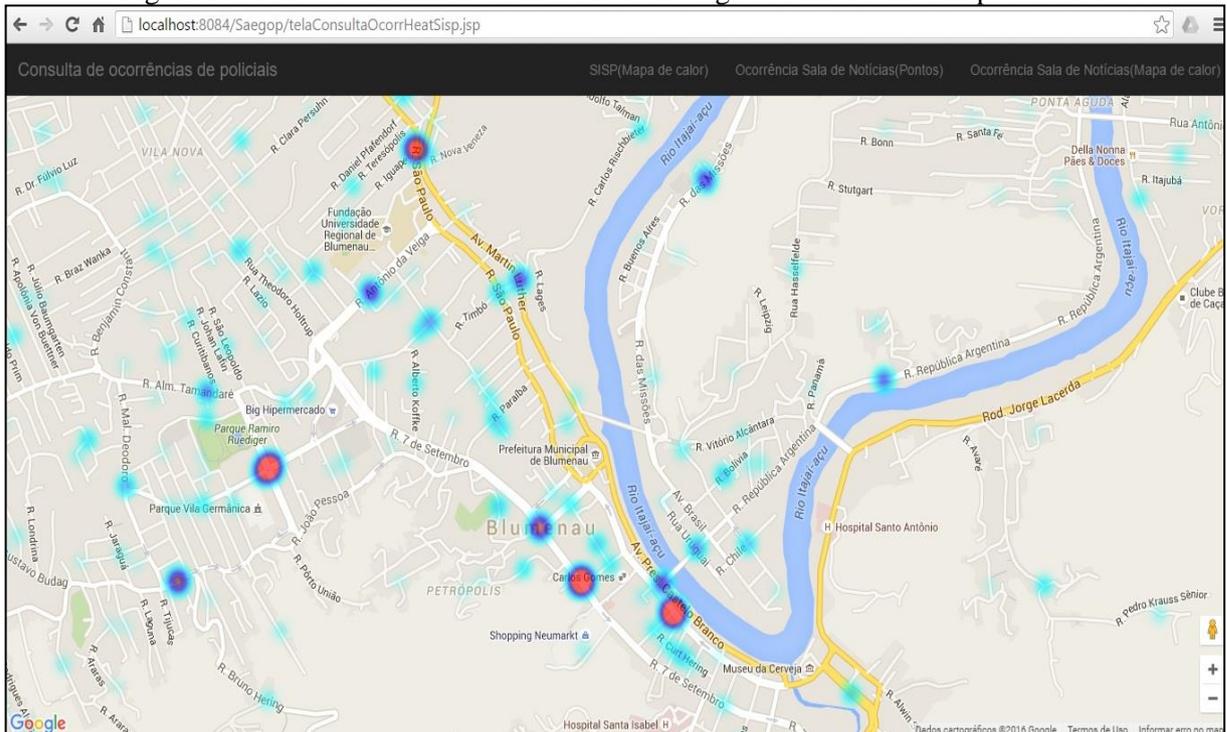
A Figura 23 apresenta o mapa de pontos do *site* de notícias com todos as ocorrências no mês de janeiro de 2016. É possível observar que o crime de furto, que está com a legenda no formato de binóculo, é o crime mais cometido no mapa. O mês de janeiro é a época do ano de período de férias, onde as pessoas acabam por realizar viagens ao litoral ou para outras localidades, deixando suas casas mais suscetíveis ao furto, por estarem desprotegidas.

Figura 23 - Ocorrências em Blumenau no mês de janeiro de 2016 com mapa de pontos



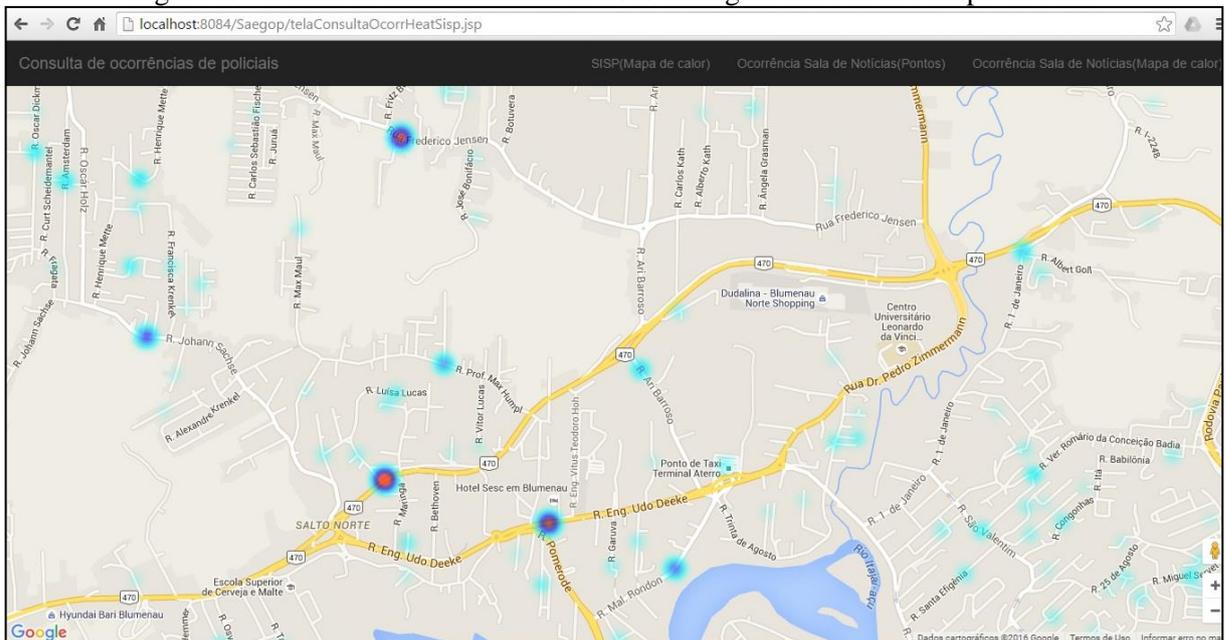
Verificando os resultados das Tabelas 1 e 2 é possível analisar que a classificação furto possui a maior quantidade de ocorrências em ambas as fontes de dados utilizadas no trabalho. Na Figura 24 é buscada todas as ocorrências de furto do ano de 2015 com os dados do relatório policial na região central da cidade de Blumenau com o mapa de calor.

Figura 24 - Furtos de Blumenau ano de 2015 na região central com o mapa de calor



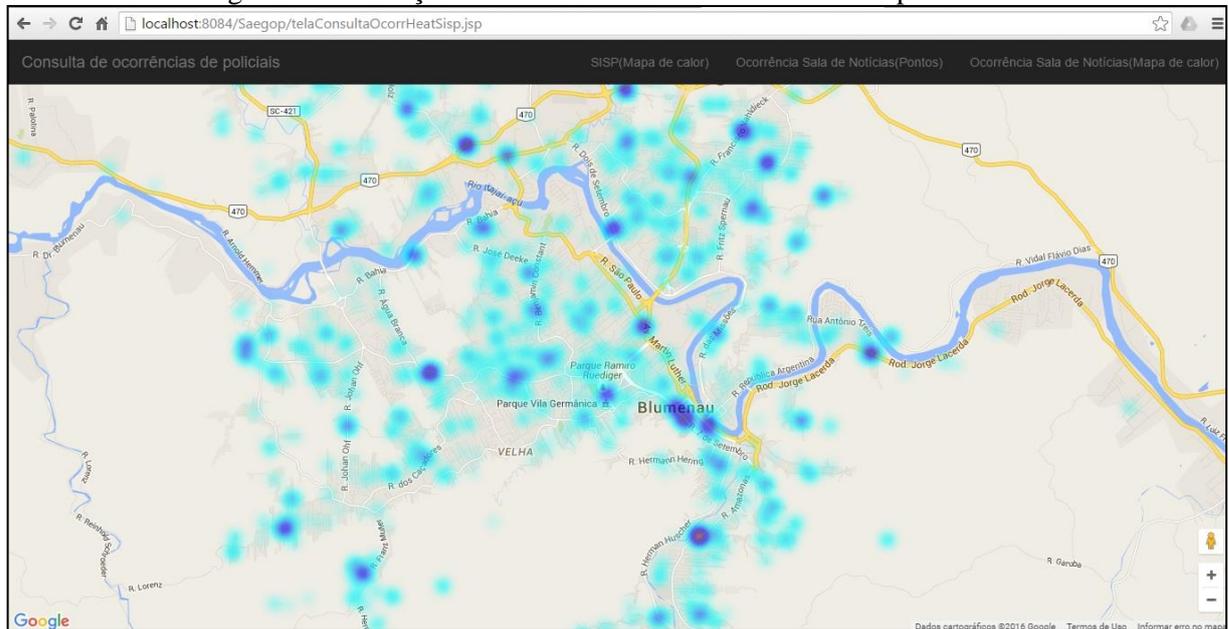
É possível notar que na região central da cidade, as ruas 7 de Setembro, São Paulo e Humberto de Campos, possui áreas em vermelhos. Isso significa que existem mais concentrações de ocorrências de furtos nessas áreas do que as demais ruas. Verificando a região norte da cidade, ruas como Engenheiro Udo Deeke, Frederico Jensen e a BR – 470 (bairro Salto do Norte), é possível analisar que também são áreas que possuem mais ocorrências de furto. A Figura 25 apresenta as ocorrências de furtos na área do norte da cidade de Blumenau com o mapa de calor.

Figura 25 - Furtos de Blumenau ano de 2015 na região norte com o mapa de calor



A Figura 26 apresenta o crime de ameaça com os dados do relatório policial na cidade de Blumenau no ano de 2015 com o mapa de calor.

Figura 26 - Ameaças em Blumenau ano de 2015 com o mapa de calor



3.4.4 Comparação de trabalhos correlatos e o sistema desenvolvido

No Quadro 14 é feita uma comparação dos trabalhos correlatos apresentados na seção 2.4 com o sistema desenvolvido. Os trabalhos que possuírem as características descritas na primeira coluna, estão marcados com “X”.

Quadro 14 - Comparação de trabalhos correlatos e sistema desenvolvido

	Bornhofen e Tefen (2009)	Furtado et al (2008)	Bordin et al (2013)	Sistema Desenvolvido (2016)
acesso <i>web</i> ao sistema	X	X		X
mapa de pontos	X		X	X
mapa de calor		X	X	X
origem da base cartográfica	GMAPS	GMAPS	Outros	GMAPS

Conforme o Quadro 14 apresenta, é possível notar que todos utilizam acesso *web* ao sistema, menos o artigo científico de Bordin et al (2013), onde não especifica como é feito o acesso ao sistema desenvolvido no artigo. Quanto a funcionalidade de mapa de pontos, apenas o trabalho de Furto et al (2008) não apresenta essa característica. O mapa de calor é presente em todos os trabalhos, exceto pelo trabalho de Bornhofen e Tefen (2009). A funcionalidade do mapa de calor ajuda na análise de criminalidade de certas regiões com maiores índices de violência. E por fim, o GMAPS é a origem da base cartográfica usado por quase todos, exceto por Bordin et al (2013), onde foi desenvolvido uma base de dados cartográfica em parceria com empresas terceiras.

4 CONCLUSÕES

O objetivo desse trabalho foi desenvolver um sistema para mapeamento geográfico de áreas de criminalidades utilizando técnicas de mineração de texto, onde conseguiu alcançar o objetivo proposto. Com os resultados obtidos é possível notar as áreas de criminalidades na região de Blumenau, se destacando os crimes de furtos nas regiões centrais e nortes da cidade.

Durante o desenvolvimento desse trabalho, foi necessário analisar diversas vezes as ocorrências coletadas através das páginas de notícias, para ter certeza que os dados extraídos estavam corretos. Como muitas páginas do *HTML* do *site* de notícias possuem problemas, foi necessário realizar algumas ações no algoritmo de coletas, para que os dados extraídos fossem buscados corretamente.

Com o processo de mineração de texto, foi possível coletar e classificar os dados não estruturados de ocorrências das páginas de notícias, onde apenas uma pequena margem de ocorrências de 5% não foram possíveis classificar, devido a estrutura do algoritmo de classificação criado. Também visto que algumas ocorrências foram classificadas de forma errada, pois alguns crimes policiais distintos eram apresentados duas vezes no texto da ocorrência.

Entretanto, como os dados de ocorrências dos relatórios que a polícia enviou já estavam pré-classificados, a abordagem de mineração de texto foi levemente utilizada, deixando um maior número de ocorrências classificadas. Ao todo, 97% das ocorrências foram classificadas.

Um dos pontos interessantes desse trabalho, foi a utilização da biblioteca PTStemmer (REUTEMANN, 2014). Com essa biblioteca, foi possível realizar a retirada do radical das palavras, tornando mais simples a classificações das ocorrências.

A *API* do GMAPS funcionou corretamente para a criação dos mapas de pontos e calor e também para buscar a geolocalização das ocorrências. Entretanto, conforme a Figura 24, é possível notar que o mapa de calor não cobre toda a rua que ocorreu o crime, fazendo-se necessário ampliar o zoom do mapa para verificar a localização correta da mancha de calor. Dessa forma, não é possível ter uma visão geral de todas as ocorrências da cidade, pois acaba sendo centralizado o zoom em um local específico.

Para concluir, outro ponto interessante foi o mapa de pontos que foi desenvolvido nesse trabalho. Com esse tipo de mapa, é possível visualizar as ocorrências por legendas de cada tipo de classificação e também ler o texto da ocorrência. Essa funcionalidade torna-se mais prático visualizar as ocorrências num período mais curto, como um mês específico e

também permite distinguir cada tipo de classificação ocorrência pela legenda apresentada no mapa.

4.1 EXTENSÕES

Algumas das extensões possíveis para este trabalho são:

- a) melhorar a coleta das ocorrências da página de notícias, para processar as ocorrências onde não foi possível buscar os fatos para a etapa de classificação;
- b) buscar outra fonte de dados relatórios policiais, que contenha o texto da ocorrência policial, como ocorre das páginas de notícias, para que possa utilizar a técnica de mineração de texto;
- c) melhorar o algoritmo de classificação, para usar outras abordagens caso não encontre a classificação da ocorrência;
- d) utilizar outras abordagens de classificação, como por exemplo a abordagem estatística, que verifica a frequência de palavras;
- e) aprimorar a busca da latitude e longitude das ocorrências, para que seja possível encontrar a geolocalização das ruas com os nomes ou as abreviaturas erradas;
- f) utilizar outra biblioteca e técnica para a criação do mapa de calor, permitindo que a rua seja marcada do início ao fim com as informações das ocorrências;
- g) criar uma tela de estatística no sistema desenvolvido, para visualizar gráficos e percentuais de ocorrências conforme um período;
- h) utilizar técnicas de associações da mineração de dados sobre os dados dos relatórios policiais, para encontrar possíveis padrões existentes entre as ocorrências policiais, de acordo com o período e classificação.

REFERÊNCIAS

- 10° BPM – MAPA DE OCORRÊNCIAS. **10° BPM – Mapa de ocorrências**. [S. l.], 2008. Disponível em: <<http://www.mapaocorrencias.site40.net/principal.php>>. Acesso em: 11 jun. 2016.
- ARANHA, C; PASSOS, E. A Tecnologia de Mineração de Textos. **Resi-Revista Eletrônica de Sistemas de Informação**, Rio de Janeiro, v.5, n. 2, p.1-8, jan. 2006.
- ARANHA, Christian N. **Uma abordagem de pré-processamento automático para mineração de textos em português**: sob o enfoque da inteligência computacional. 2007. 144f. Tese (Doutorado em Engenharia Elétrica) – Programa de Pós-Graduação em Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro. Disponível em: <http://www.maxwell.lambda.ele.puc-rio.br/10081/10081_4.pdf>. Acesso em: 04 set. 2015.
- BEATO, C. C. Determinantes da criminalidade em Minas Gerais. **Revistas Brasileiras de Ciências Sociais**, Minas Gerais, v. 13, n. 37, p.1-20, jun. 1998.
- BEPPLER, M. D.; FERNANDES, A. M. da R. Aplicação de Text Mining para a extração de conhecimento jurisprudencial. In: CONGRESSO SUL BRASILEIRO DE COMPUTAÇÃO, 2005, Criciúma. **Anais...** Criciúma: I Congresso Sul Brasileiro de Computação, 2005. Não paginado.
- BLOG DO VASCO. **Blog do Vasco**. [S. l.], 2011. Disponível em: <<http://vfurtado.blogspot.com.br/2011/03/wikicrimes-com-novidades.html>>. Acesso em: 11 jun. 2016.
- BORDIN, M., et al. O uso do geoprocessamento na segurança pública do Estado do Paraná. In: XVI SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 16., 2013, Foz do Iguaçu. **Anais...** Foz do Iguaçu: SBSR, 2013. Disponível em: <<http://www.dsr.inpe.br/sbsr2013/files/p0166.pdf>>. Acesso em: 11 jun. 2016.
- BORGES, D. Vitimização e Sentimento de Insegurança no Brasil em 2010: Teoria, Análise e Contexto. **Mediações-Revista de Ciências Sociais**, Londrina, v. 18, n. 1, p. 141-163, jun. 2013.
- BORNHOFEN, P. R.; TENFEN, E. Mapeamento criminal por meio da plataforma Google Maps. **Revista Brasileira de Segurança Pública**, v. 5, n. 3, p.82-98, 01 set. 2009.
- BOOTSTRAP. **Documentação da ferramenta Bootstrap**. [S. l.], 2016. Disponível em: <<http://getbootstrap.com/css/>>. Acesso em: 11 jun. 2016.
- BRASIL. **Lei 12.527/2011**. Brasília, Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/12527.htm>. Acesso em: 11 jun. 2016.
- CAMARGO, O. **Violência no Brasil, outro olhar**. São Paulo, 2007. Disponível em <<http://www.brasilecola.com/sociologia/violencia-no-brasil.htm>>. Acesso em: 11 de jun. 2016.
- CAMILO, C. O.; SILVA, J. C. **Mineração de Dados: Conceitos, Tarefas, Métodos e Ferramentas**. Goiânia, 2009. Disponível em: <http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_001-09.pdf>. Acesso em: 17 jun. 2016.
- CLEG, H. **Wiki pinpoints Brazilian crime**. Londres, 2008. Disponível em: <<http://news.bbc.co.uk/2/hi/technology/7347101.stm>>. Acesso em: 11 de jun. 2016.

COSME, A. **Projeto em sistemas de informação geográfica**. Lisboa: Lidel – Edição Técnicas, 2012.

FERREIRA, I. F. C. B.; PENNA, N. A. Território da violência: um olhar geográfico sobre a violência urbana. **GEOUSP: espaço e tempo**, São Paulo, v. 1, n. 18, p.155-168, jan. 2005.

FURTADO, V. et al. WikiCrimes - Um Sistema Colaborativo para Mapeamento Criminal. In: CONGRESSO TECNOLÓGICO INFOBRASIL TI & TELECOM, 1., 2008, Fortaleza. **Anais...**. Fortaleza: Infobrasil, 2008. p. 1 - 6. Disponível em: <[http://www.infobrasil.inf.br/userfiles/WikiCrimes_Um Sistema Colaborativo para Mapeamento Criminal.pdf](http://www.infobrasil.inf.br/userfiles/WikiCrimes_Um%20Sistema%20Colaborativo%20para%20Mapeamento%20Criminal.pdf)>. Acesso em: 11 jun. 2016.

GOOGLE DEVELOPERS. **Documentação de desenvolvedor do Google Maps**. [S. l.], 2016. Disponível em: <<https://developers.google.com/maps/documentation/?hl=pt-br>>. Acesso em: 11 jun. 2016.

GOOGLE MAPS. **Google Maps**. [S. l.], 2016. Disponível em: <<https://www.google.com.br/maps/@-26.9136301,-49.0698426,14z>>. Acesso em: 11 jun. 2016.

LEMLE, M. **Blumenau tem mapa do crime atualizado todo dia na internet**. [S. l.], 2009. Disponível em: <<http://www.comunidadesegura.org.br/MATERIA-mapas-interativos-informam-populacao-sobre-crimes>>. Acesso em: 11 jun. 2016.

LOBO, M. A. A.; GUIMARÃES, L. H. R. Distribuição espacial da criminalidade no Centro Histórico da cidade de Belém (Pará/Brasil). **Revista Electrónica de Geografía y Ciencias Sociales**, Barcelona, Espanha, v. 17, n. 456, p. 1-16, mai. 2013.

JORNAL DE SANTA CATARINA. **O MAPA DOS FURTOS EM BLUMENAU**. Blumenau, 2016. Disponível em: <<http://jornaldesantacatarina.clicrbs.com.br/sc/geral/pagina/o-mapa-dos-furtos-em-blumenau.html>>. Acesso em: 11 jun. 2016.

JSOUPORG. **jsoup: Java HTML Parser**. [S. l.], 2009. Disponível em: <<https://jsoup.org/>>. Acesso em: 11 jun. 2016.

METROPOLITAN POLICE. **Metropolitan Police Crime Mapping**. Londres, 2016. Disponível em: <<http://maps.met.police.uk/>>. Acesso em: 11 jun. 2016.

MORICONI, L. H. **Mapas interativos informam população sobre crimes**. [S. l.], 2009. Disponível em: <<http://www.comunidadesegura.org.br/MATERIA-mapas-interativos-informam-populacao-sobre-crimes>>. Acesso em: 11 jun. 2016.

MORAIS, E. A. M.; AMBRÓSIO, A. P. L. **Mineração de Textos**. Goiânia, 2007. Disponível em: <http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_005-07.pdf>. Acesso em: 11 jun. 2016.

PAULA, de G.; DANDOLINI, G. A.; SOUZA, J. A. Tecnologia da informação e comunicação e as atividades de inteligência. **Revista de Ordem Pública**, Florianópolis, v. 5, n.1, p. 119-137, jan. 2012.

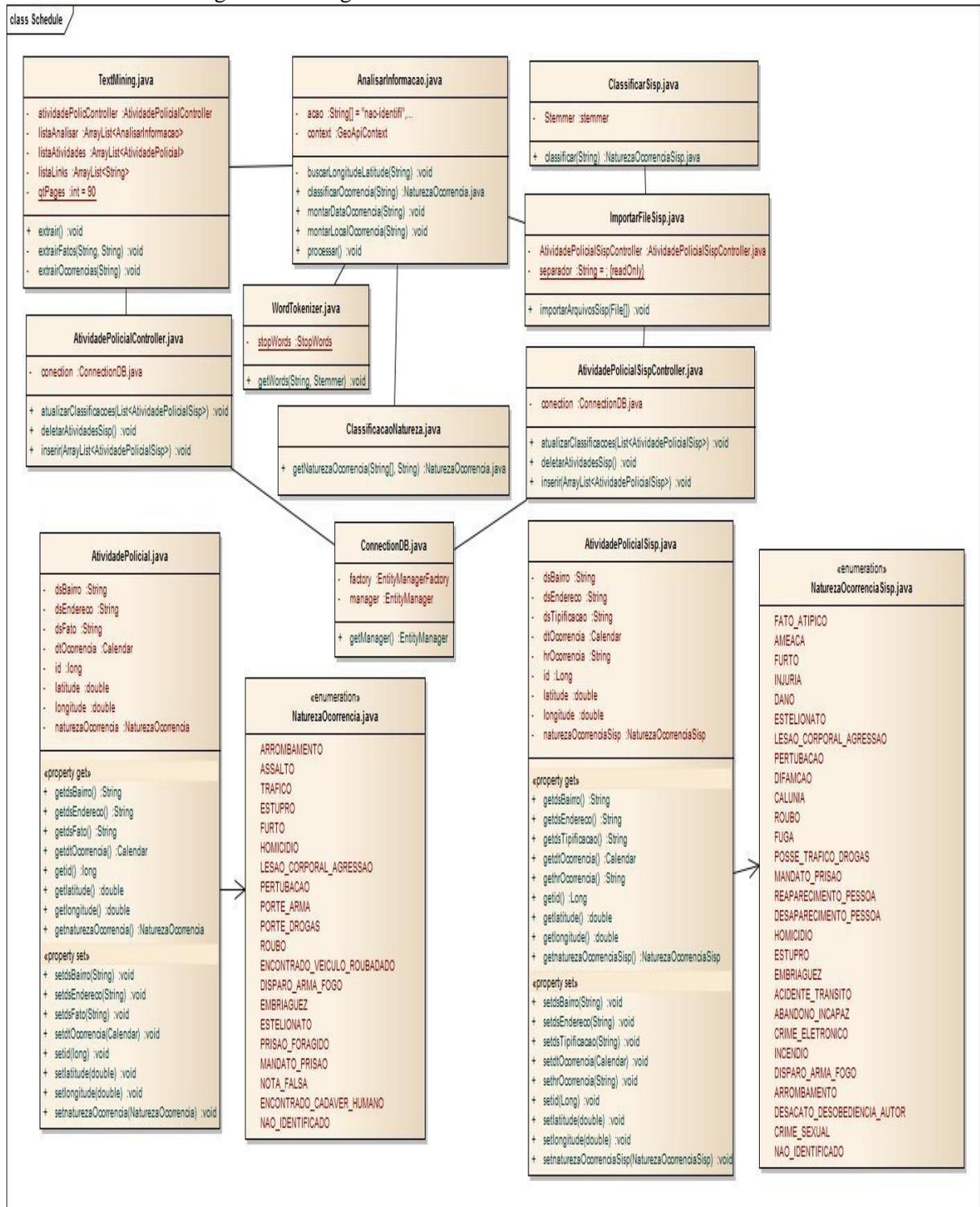
QGIS TUTORIALS AND TIPS. **Criando Mapas de Calor**. [S. l.], 2015. Disponível em: <http://www.qgistutorials.com/pt_BR/docs/creating_heatmaps.html>. Acesso em: 11 jun. 2016.

- QUEIROZ, J. W. de. **Sistemas de informações geográfica e análise espacial de dados como ferramentas para determinação de agregação espacial de doenças: A Hanseníase como modelo.** 2009. 161f. Tese (Doutorado em Ciências da Saúde) - Programa de Pós-Graduação em Ciências da Saúde, Universidade Federal do Rio Grande do Norte, Natal.
- REUTEMANN, P. **Weka package for the PTStemmer.** . [S. l.], 2014. Disponível em: <<https://github.com/fracpete/ptstemmer-weka-package>>. Acesso em: 15 jun. 2016.
- REZENDE, S. O. **Sistemas inteligentes: fundamentos e aplicações.** Barueri: Editora Manole Ltda, 2003. 525 p, il.
- RODRIGUES, V. **Protótipo de um sistema de análise espacial de dados em redes sociais.** 2014. 77f. Trabalho de Conclusão de Curso (Bacharelado em Sistema de Informação) – Unidade Acadêmica de Graduação, Universidade do Vale do Rio dos Sinos, São Leopoldo.
- SABINO, A. R. **Ferramenta para criação de bases de conhecimento na forma de ontologia OWL a partir de dados não estruturados.** 2014. 117f. Trabalho de Conclusão de Curso (Bacharelado em Ciências da Computação) – Centro de Ciências Exatas e Naturais, Universidade Regional de Blumenau, Blumenau. Disponível em: <http://dsc.inf.furb.br/arquivos/tccs/monografias/2014_1_allan-renato_sabino_monografia.pdf>. Acesso em: 04 set. 2015.
- SALA DE NOTÍCIAS. **SALA DE NOTÍCIAS.** [S. l.], 2012. Disponível em: <<http://www.saladenoticias.net/?s=atividade%20operacionais&submit=Pesquisar>>. Acesso em: 11 jun. 2016.
- SANTANA, L. L. S.; SILVA, M. P.; CHAGAS, Clay A. N. O uso da produção cartográfica na análise do território e da violência no bairro do Guamá em Belém-PA. **Revista Geonorte, Edição Especial**, Manaus, v. 7, n. 1, p. 1623-1636, 2013.
- SANTOS, F. H. M., et al. Use of Computational Tools in the Support to Accomplishment of Criminal Analysis in the Public Security. In: 4TH CONTECSI INTERNATIONAL CONFERENCE ON INFORMATION SYSTEMS AND TECHNOLOGY MANAGEMENT., 4, 2007, São Paulo. **Anais...** São Paulo: CONTECSI, 2007. p. 1914-1926.
- SCUSSEL, A. **Artigo: Por dentro do Google Maps,** [S. l.], 2015. Disponível em: <<http://mundogeo.com/blog/2013/07/01/artigo-por-dentro-do-google-maps/>>. Acesso em: 11 jun. 2016.
- SOARES, F. de A. **Categorização automática de textos baseada em mineração de texto.** 2013. 158 f. Tese (Doutorado em Engenharia Elétrica) – Programa de Pós-Graduação em Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro. Disponível em: < http://www.maxwell.vrac.puc-rio.br/23213/23213_3.PDF>. Acesso em: 04 set. 2015.
- WIVES, L. **Recursos de text mining.** Porto Alegre, 2000. Disponível em: <<http://www.inf.ufrgs.br/~wives/english/textmining.html> >. Acesso em: 11 jun. 2016.

APÊNDICE A – Diagrama da classe Schedule com atributos e métodos

Este apêndice apresenta (Figura 27) o diagrama de classe do Schedule com os atributos e métodos.

Figura 27 - Diagrama classe Schedule com atributos e métodos



APÊNDICE B – Lista de *stopwords* removida pelo algoritmo

Este apêndice apresenta (Quadro 15) a lista de *stopwords* que foram removidas pelo algoritmo, conforme a técnica de pré-processamento da MT.

Quadro 15 - Lista de *StopWords*

<p>O, a, os, as, um, uma, uns, umas, eu, tu, ele, ela, nós, vós, eles, elas, me, mim, comigo, te, ti, contigo, se, si, lhe, o, a, nos, conosco, vos, lhes, os, as, meu, minha, meus, minhas, teu, tua, teus, tuas, seu, sua, seus, suas, nosso, nossa, nossos, nossas, vosso, vossa, vossos, vossas, este, estes, esta, estas, esse, esses, essa, essas, aquele, aqueles, aquela, aquelas, mesmo, mesmos, mesma, mesmas, próprio, próprios, própria, próprias, tal, tais, isto, isso, aquilo, sim, deveras, talvez, acaso, porventura, decerto, muito, pouco, assaz, bastante, mais, menos, tão, demasiado, meio, todo, demais, que, quão, quanto, quase, como, abaixo, acima, cá, lá, aqui, ali, aé, além, aquém, algures, alhures, nenhures, atrás, fora, afora, dentro, perto, longe, adiante, diante, onde, avante, através, defronte, aonde, donde, bem, mal, assim, depressa, devagar, de balde, alerta, melhor, pior, não, tampouco, agora, hoje, amanhã, depois, ontem, anteontem, já, sempre, nunca, jamais, ainda, logo, antes, cedo, tarde, ora, outrora, então, absolutamente, breve, calmamente, certamente, corretamente, hereupon, hers, herself, him, himself, his, how, efetivamente, fielmente, levemente, possivelmente, primeiramente, provavelmente, realmente, tanto, tarde, cujo, cujos, cuja, cujas, quanto, quantos, quanta, quantas, quem, eis, exclusive, menos, exceto, fora, salvo, senão, sequer, inclusive, também, mesmo, ainda, até, ademais, além disso, de mais a mais, sã, apenas, sobretudo, embora, aliás, ou, melhor, isto é, ou antes, a saber, por exemplo, ou seja, afinal, do, da, policia, militar, fato:, no, na, mas, ocorrencia, porta, portas</p>

APÊNDICE C – Lista de tipos de classificações das ocorrências

Este apêndice apresenta as listas de classificações que foram utilizadas nas ocorrências das páginas de notícias e dos relatórios policiais.

O Quadro 16 apresenta as classificações das ocorrências das páginas de notícias.

Quadro 16 - Lista de classificações das ocorrências das páginas de notícias

Arrombamento, Assalto, Tráfico, Estupro, Furto, Homicídio, Porte de Arma, Estelionato, Lesão Corporal/Agressão, Perturbação, Porte de Drogas, Roubo, Embriaguez, Encontrado veículo roubado, Disparo de arma de fogo, Pagamento com moeda falsa, Prisão de foragido, Mandato de prisão, Encontrado cadáver humano, Não identificado.

O Quadro 17 apresenta as classificações das ocorrências dos relatórios policiais.

Quadro 17 - Lista de classificações das ocorrências dos relatórios policiais

Fato atípico, Ameaça, Furto, Injúria, Dano, Estelionato, Difamação, Perturbação, Roubo, Lesão Corporal/Agressão, Calúnia, Fuga, Posse/Tráfico de drogas, Mandato de prisão, Reaparecimento de pessoa, Desaparecimento de pessoa, Homicídio/Suicídios/Mortes adversas, Estupro, Embriaguez, Acidente de trânsito, Abandono de incapaz, Crimes eletrônicos, Incêndio, Arrombamento, Crime sexual, Disparo/Porte de arma de fogo, Desacato/Desobediência pela autoridade, Não identificado.

APÊNDICE D – Listas de palavras para classificações das ocorrências

Este apêndice apresenta as listas de palavras que foram utilizadas para classificar as ocorrências coletadas das páginas de notícias e dos relatórios policiais. As palavras dessas listas estão sem o radical.

O Quadro 18 apresenta a lista de verbos utilizados para classificação de ocorrências nas páginas de notícias.

Quadro 18 - Verbos para classificações

"foi", "auxili", "atend", "prend", "apreend", "efetu", "registr", "prest", "denunc", "receb", "recuper", "abord"

O Quadro 19 apresenta os crimes policiais utilizados para classificação de ocorrências nas páginas de notícias.

Quadro 19 - Crimes policiais para classificação páginas de notícias

"furt", "furto", "assalt", "roub", "perturb", "traf", "agred", "mand", "arromb", "homicidi", "embriagu", "agressa", "port", "forag", "perturbaca", "dispar", "encontr", "estupr", "fer", "desfer", "corpor", "estelionat", "esfaque", "not", "possu", "substanc", "poss"

O Quadro 20 apresenta os objetos policiais utilizados para classificação de ocorrências nas páginas de notícias.

Quadro 20 - Objetos policiais para classificações

"maconh", "drog", "crack", "arm", "cocain", "veicul", "mot", "motociclet", "pris", "prisa", "apreensa", "pistol", "revolv", "pedr", "entorpec", "cadav", "fals", "embriag", "obit", "carr"

O Quadro 21 apresenta os crimes policiais utilizados para classificação de ocorrências dos relatórios policiais.

Quadro 21 - Crimes policiais para classificação dos relatórios policiais

"atip", "ameac", "furt", "injur", "dan", "via", "estelionat", "corpor", "perturb", "perturbaca", "difamaca", "calun", "caluni", "roub", "fug", "poss", "traf", "mand", "reaparec", "eletron", "desaparec", "incendi", "estupr", "mort", "fraud", "port", "falsificaca", "falsidad", "abandon", "dispar", "suicidi", "vand", "latrocini", "embriagu", "ilicit", "acid", "violenc", "dirig", "rac", "invasa", "homofob", "comunicaca", "apropriaca", "trat", "desacat", "desobedienc", "libidin", "obscen", "sex", "prostituica", "constrang", "homicidi"

O Quadro 22 apresenta os crimes policiais que precisam de um objeto policial para ser classificado.

Quadro 22 - Crimes policiais que precisam de objeto policial

"encontr", "poss", "possu", "substanc", "mand", "not", "port"
