

Gabriel da Silva Almeida

Um modelo computacional para mineração de dados no Facebook aplicado a inferência dos indicadores de cidades inteligentes

Rio Grande, RS, Brasil

Setembro de 2018

MINISTÉRIO DE EDUCAÇÃO
UNIVERSIDADE FEDERAL DO RIO GRANDE
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM COMPUTACIONAL

**Um modelo computacional para mineração de dados
no Facebook aplicado a inferência dos indicadores de
cidades inteligentes**

Por
Gabriel da Silva Almeida

Dissertação para obtenção do Título de Mestre em Modelagem Computacional

Orientadora: Prof^a. Dr^a. Silvia Silva da Costa Botelho

Rio Grande, RS, Brasil

Setembro de 2018

Ficha catalográfica

A447m Almeida, Gabriel da Silva.

Um modelo computacional para mineração de dados no Facebook aplicado a inferência dos indicadores de cidades inteligentes / Gabriel da Silva Almeida. – 2018.

175f.

Dissertação (mestrado) – Universidade Federal do Rio Grande – FURG, Programa de Pós-Graduação em Modelagem Computacional, Rio Grande/RN, 2018.

Orientadora: Dra. Silvia Silva da Costa Botelho.

1. Redes Sociais 2. Facebook 3. Cidades Inteligentes
4. Mineração de Textos 5. Mineração de Dados I. Botelho, Silvia Silva da Costa II. Título.

CDU 004.414.23

Gabriel da Silva Almeida

“Um modelo computacional para mineração de dados no Facebook aplicado a inferência dos indicadores de cidades inteligentes”

Dissertação apresentada ao Programa de Pós Graduação em Modelagem Computacional da Universidade Federal do Rio Grande - FURG, como requisito parcial para obtenção do Grau de Mestre. Área concentração: Modelagem Computacional.

Aprovado em

BANCA EXAMINADORA

Prof^a. Dr^a. Silvia Silva da Costa Botelho
Orientadora – FURG

Prof. Dr. Emanuel da Silva Diaz Estrada
FURG

Prof. Dr. Rafael Augusto Penna dos Santos
FURG

Prof. Dr. Eduardo Wenzel Brião
IFRS – Rio Grande

Rio Grande - RS
2018

AGRADECIMENTOS

Agradeço primeiramente à Deus por estar vivo, pois estar vivo é a resposta de Deus a todos os nossos problemas; à minha família, por ter sempre me prestado o suporte e apoio para que eu estudasse, à Universidade Federal do Rio Grande pelo ensino de qualidade, à minha orientadora Prof^a. Dr^a. Silvia Silva da Costa Botelho por ter me orientado neste trabalho, à CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível) pelo suporte financeiro para realização desta pesquisa e a todos que de alguma forma contribuíram para meu crescimento e aprendizado.

“A sabedoria dos homens é proporcional não à sua experiência mas à sua capacidade de adquirir experiência”.

George Bernard Shaw

RESUMO

Nos últimos anos, as pesquisas científicas têm voltado o foco para dados armazenados em meio digital sem a preocupação com o rigor da estruturação. No mundo em que vivemos, cada vez mais pessoas têm acesso sem restrições à Internet e podem divulgar suas ideias através das redes sociais. Isto faz com que exista um crescente volume de textos espalhados pela Web, principalmente nas redes sociais. Publicações em redes sociais contêm informações que, muitas vezes, podem expressar percepções relevantes a serem consideradas para a tomada de decisão, por exemplo, na gestão pública. O modelo de cidade inteligente vem com a proposta de monitorar e integrar as condições de operações das infraestruturas críticas da cidade, atuando de forma preventiva para a continuidade de suas atividades essenciais, melhorando as condições de serviços e a qualidade de vida dos cidadãos. Diante deste cenário, este trabalho utiliza técnicas de mineração de dados para serem aplicadas a coleções de postagens em redes sociais, de forma a inferir o entendimento da população a respeito de um determinado assunto ou tema. Mais precisamente, no âmbito de cidades inteligentes e seus indicadores, o objetivo desta dissertação é propor o uso de técnicas de classificação baseadas em termos representativos para que a partir destes possam ser extraídos mapas conceituais sobre os principais indicadores que compõem o nível de inteligência de um município. Com este estudo pretende-se analisar dados expressos pela população em meios às redes sociais sobre diferentes aspectos e gerar resultados que poderão ser utilizados por agentes públicos para auxiliar à tomada de decisão. Os aspectos tratam-se de características sobre determinada entidade, os quais são anotados para o processo de mineração. Além de avaliar conceitos relacionados à mineração de textos, este estudo foi aplicado em dados de grupos e páginas do Facebook que contêm informações relacionadas à cidade do Rio Grande/RS, com o intuito de classificar os textos por aspectos. Ademais, essa investigação assume o desafio tecnológico de realizar uma mineração de textos em fontes fracamente estruturadas, considerando que os dados textuais contêm pouca ou quase nenhuma estrutura. Os resultados obtidos neste estudo demonstram que o método pode ser aplicado, uma vez que as métricas de avaliação foram adequadas. A partir dos termos representativos e das classificações obtidas, é possível observar indícios sobre a percepção coletiva da população relacionada aos indicadores de qualidade de vida no âmbito de cidades inteligentes.

Palavras-chaves: Redes Sociais, Facebook, Cidades Inteligentes, Mineração de Textos, Mineração de Dados

ABSTRACT

In recent years, scientific research has focused on data stored in digital media without concern for the rigor of structuring. In the world we live in, more and more people have unrestricted access to the Internet and can spread their ideas through social networks. This causes a growing volume of texts spread throughout the Web, especially in social networks. Publications in social networks contain information that can often express relevant perceptions to be considered for decision making, for example, in public management. The intelligent city model comes with the proposal to monitor and integrate the operational conditions of the city's critical infrastructures, acting in a preventive way for the continuity of its essential activities, improving the conditions of services and the quality of life of citizens. Given this scenario, this work uses data mining techniques to be applied to collections of postings in social networks, in order to infer the understanding of the population about a particular subject or theme. More precisely, in the context of intelligent cities and their indicators, the objective of this dissertation is to propose the use of classification techniques based on representative terms so that from these can be extracted conceptual maps on the main indicators that make up the level of intelligence of a County. This study intends to analyze data expressed by the population in social networks media on different aspects and generate results that can be used by public agents to assist in decision making. The aspects are about characteristics about a particular entity, which are annotated for the mining process. In addition to evaluating concepts related to text mining, this study was applied to data from groups and Facebook pages that contain information related to the city of Rio Grande / RS, in order to classify the texts by aspects. In addition, this research assumes the technological challenge of performing a text mining in weakly structured sources, considering that textual data contain little or no structure. The results obtained in this study demonstrate that the method can be applied, since the evaluation metrics were adequate. From the representative terms and the classifications obtained, it is possible to observe clues about the collective perception of the population related to the indicators of quality of life within the framework of intelligent cities.

Keywords: Social Networks, Facebook, Smart Cities, Text Mining, Data Mining

SUMÁRIO

1 INTRODUÇÃO	27
1.1 Motivação e Justificativa	29
1.2 Objetivos	31
1.2.1 Objetivo Geral	31
1.2.2 Objetivos Específicos	32
1.3 Organização	32
2 FUNDAMENTAÇÃO TEÓRICA	33
2.1 Cidades Inteligentes	33
2.2 Indicadores para Cidades Inteligentes	33
2.3 Mineração de Dados	36
2.4 Mineração de Textos	40
2.4.1 Seleção dos Textos e Tipos de Abordagens de Dados	41
2.4.2 Preparação dos Dados	42
2.4.3 Indexação e Normalização	43
2.4.4 Cálculo de Relevância dos Termos	45
2.4.5 Seleção dos Termos	46
2.4.6 Pós-Processamento ou Análise de Resultados	47
2.4.7 Técnicas de Análise de Coocorrência entre Termos em Documentos	47
2.4.8 Expected Mutual Information Measure	48
2.4.9 Phi-squared	49
3 TRABALHOS RELACIONADOS	51
3.1 Estado da Arte	51
3.1.1 Monitoramento de Mídias Sociais no Processo de Análise da Percepção dos Usuários Sobre a Segurança de Porto Alegre	51
3.1.2 Utilizando Redes Sociais para Detectar Crimes em Tempo Real	52
3.1.3 Análise de Redes Sociais com Uso de Aprendizado de Máquina para Prever o Tráfego de Veículos em Zonas Urbanas	53

3.1.4	Uma Ferramenta de Rede Social Baseada em Localização Inspirada em Soluções de Cidades Inteligentes	57
3.1.5	Priorização de Nós de uma Rede de Sensores sem Fio com Base na Análise de Dados de Redes Sociais	58
3.2	Análise comparativa	60
4	UM MODELO PARA MINERAÇÃO DE DADOS DE POSTAGENS DO FACEBOOK PARA INFERÊNCIA DA INTELIGÊNCIA DO TERRITÓRIO	63
4.1	Geração de documentos de domínio	66
4.2	Pré-processamento	68
4.2.1	Seleção dos textos	68
4.2.2	Geração do conjunto de treinamento	68
4.2.3	Remoção de <i>stop words</i> e termos simples	70
4.2.4	Cálculos de frequência dos termos	71
4.2.5	Anotação de termos substantivos	72
4.2.6	Preparação dos dados	73
4.3	Palavras Sementes	73
4.4	Geração de Palavras Candidatas	73
4.5	Geração dos Termos Representativos	74
4.6	Identificação de aspectos nos documentos	74
5	TESTES E VALIDAÇÃO	77
5.1	Delimitação do Estudo de Caso	77
5.2	Anotação dos dados	80
5.3	Aquisição e pré-processamento dos dados anotados	80
5.4	Geração das Palavras Candidatas nos dados anotados	81
5.5	Definição das Palavras Sementes	81
5.6	Geração de Termos Representativos e Classificação dos Textos nos dados anotados	83
5.7	Classificação e geração dos termos representativos nos dados não anotados	85
6	CONSIDERAÇÕES FINAIS	89
REFERÊNCIAS		91

APÊNDICE A TELAS DO SISTEMA WEB PROPOSTO PARA ANOTAÇÃO DOS DOCUMENTOS	95
APÊNDICE B STOP WORDS E TERMOS SIMPLES ANOTADOS	99
APÊNDICE C PALAVRAS SEMENTES ANOTADAS PARA CADA ASPECTO	101
APÊNDICE D EXPERIMENTOS REALIZADOS	109
APÊNDICE E TERMOS REPRESENTATIVOS A PARTIR DO DATASET ANOTADO	121
APÊNDICE F TERMOS REPRESENTATIVOS MAIS FREQUENTES A PARTIR DO DATASET ANOTADO	141
APÊNDICE G TERMOS REPRESENTATIVOS A PARTIR DE TODOS DOCUMENTOS CLASSIFICADOS	151
APÊNDICE H TERMOS REPRESENTATIVOS MAIS FREQUENTES A PARTIR DE TODOS DOCUMENTOS CLASSIFICADOS	163
ANEXO A TABELAS DE INDICADORES DE CIDADES INTELIGENTES	173

LISTA DE FIGURAS

Figura 1.1 - Número de usuários de mídias sociais em todo o mundo no período de 2010 a 2021 (em bilhões). Fonte: (STATISTA, 2017a).	30
Figura 1.2 - Número de usuários do Facebook no Brasil de 2015 a 2018 (em milhões). Fonte: (STATISTA, 2017b).	30
Figura 1.3 - Redes sociais mais populares no Brasil a partir de julho de 2015, com base no número de visitantes únicos (em milhões). Fonte: (STATISTA, 2017c).	31
Figura 2.1 - Mineração de dados e técnicas de diversas áreas relacionadas. Fonte adaptada de Han, Pei e Kamber (2011).	36
Figura 2.2 - Abordagens utilizadas no processo de Descoberta de Conhecimento. Fonte adaptada de Morais e Ambrósio (2007).	37
Figura 2.3 - Visão geral das etapas do processo de Descoberta de Conhecimento em Base de Dados - KDD. Fonte adaptada de Passos e Goldschmidt (2005).	38
Figura 2.4 - Etapas do processo de mineração de textos. Fonte: Morais e Ambrósio (2007).	42
Figura 3.1 - Plataforma para detecção de crimes em tempo real. Fonte: (FIGUEREDO; SOUSA, 2016).	52
Figura 3.2 - <i>Heatmap</i> dos incidentes que ocorreram às 07:43:00 no dia 12/06/2017 - Segunda-Feira. Fonte: (DIAS, 2017).	56
Figura 3.3 - Mapa com marcadores dos Incidentes das classes Acidente e Congestionamento às 17:42:04 no dia 12/06/2017 - Segunda-Feira. Com um marcador de congestionamento selecionado. Fonte: (DIAS, 2017).	57
Figura 3.4 - Esquema geral da abordagem proposta. Fonte: (ANDRADE, 2017).	59
Figura 3.5 - Eventos detectados no <i>dataset</i> de Nova York. Fonte: (ANDRADE, 2017).	60
Figura 4.1 - Fluxograma da abordagem proposta.	64
Figura 4.2 - Fluxograma da abordagem proposta.	65
Figura 4.3 - Diagrama do modelo objeto-relacional da base de dados.	67

Figura 4.4 - Tabela <i>DocumentClassification</i> para armazenamento das anotações dos documentos e tabelas relacionadas <i>Document</i> e <i>UserClassificationWebSystem</i>	69
Figura 4.5 - Tabelas criadas e populadas para armazenar os documentos anotados nos respectivos aspectos, após julgados por unânimidade ou maioria absoluta de votos no aspecto.	70
Figura 4.6 - Tabelas <i>WordCloud</i> e <i>WordCloudTF</i> da base de dados MySQL, relacionadas à tabela <i>Document</i> , respectivamente para geração de um modelo booleano entre termos e documentos e armazenamento das frequências dos termos nos documentos.	72
Figura 4.7 - Tabelas EMIMPhiSquared para armazenamento de termos representativos.	75
Figura 5.1 - Nuvem de palavras para o aspecto Segurança.	86
Figura 5.2 - Nuvem de palavras para o aspecto Segurança.	87
Figura A.1 - Tela de criação de usuário do sistema web proposto.	95
Figura A.2 - Tela de login do sistema web proposto.	95
Figura A.3 - Tela de apresentação das publicações no sistema web proposto. Parte 1.	96
Figura A.4 - Tela de apresentação das publicações no sistema web proposto. Parte 2.	96
Figura A.5 - Tela de apresentação das publicações no sistema web proposto. Parte 3.	97
Figura A.6 - <i>Query</i> para apresentação dos documentos no sistema web proposto.	97
Figura A.7 - <i>Query SQL</i> para a criação da tabela <i>WordCloudTF</i>	98
Figura A.8 - <i>Query SQL</i> para a população da tabela <i>WordCloudTF</i> a partir dos registros da tabela <i>WordCloud</i>	98
Figura F.1 - Nuvem de palavras para o aspecto Economia.	141
Figura F.2 - Nuvem de palavras para o aspecto Educação.	142
Figura F.3 - Nuvem de palavras para o aspecto Empreendedorismo.	143
Figura F.4 - Nuvem de palavras para o aspecto Energia.	144
Figura F.5 - Nuvem de palavras para o aspecto Governança.	145
Figura F.6 - Nuvem de palavras para o aspecto Meio Ambiente.	146
Figura F.7 - Nuvem de palavras para o aspecto Mobilidade.	147
Figura F.8 - Nuvem de palavras para o aspecto Saúde.	148

Figura F.9 - Nuvem de palavras para o aspecto Tecnologia & Inovação.	149
Figura F.10 - Nuvem de palavras para o aspecto Urbanismo.	150
Figura H.1 - Nuvem de palavras para o aspecto Economia.	163
Figura H.2 - Nuvem de palavras para o aspecto Educação.	164
Figura H.3 - Nuvem de palavras para o aspecto Empreendedorismo.	165
Figura H.4 - Nuvem de palavras para o aspecto Energia.	166
Figura H.5 - Nuvem de palavras para o aspecto Governança.	167
Figura H.6 - Nuvem de palavras para o aspecto Meio Ambiente.	168
Figura H.7 - Nuvem de palavras para o aspecto Mobilidade.	169
Figura H.8 - Nuvem de palavras para o aspecto Saúde.	170
Figura H.9 - Nuvem de palavras para o aspecto Urbanismo.	171

LISTA DE TABELAS

Tabela 2.1 - Análise comparativa dos Rankings de Inteligência Urbana	35
Tabela 3.1 - Tabela Comparativa dos Trabalhos Relacionados.	61
Tabela 5.1 - Grupos da rede social Facebook, relacionados às informações da cidade do Rio Grande/RS.	78
Tabela 5.2 - Continuação da Tabela 5.1. Grupos da rede social Facebook, relacionados às informações da cidade do Rio Grande/RS.	79
Tabela 5.3 - Quantidade de documentos anotados em cada aspecto a partir do sistema web.	80
Tabela 5.4 - Quantidade de Palavras Candidatas geradas por cada experimento	81
Tabela 5.5 - Palavras Sementes para Segurança.	82
Tabela 5.6 - Experimentos selecionados para cada aspecto	83
Tabela 5.7 - Experimentos selecionados para cada aspecto	84
Tabela 5.8 - Resultados do classificador final referente à matriz de confusão para os aspectos de estudo deste trabalho.	85
Tabela 5.9 - Quantidade de documentos classificados por aspecto, considerando os respectivos termos representativos gerados e selecionados, apresentados nas Tabelas E.1, E.2, E.3, E.4, E.5, E.6, E.7, E.8, E.9, E.10, E.11, E.12, E.13, E.14, E.16, E.17, E.18, E.19, E.20 e E.21, localizadas no apêndice deste documento.	86
Tabela B.1 - <i>Stop Words</i> e alguns termos simples anotados.	100
Tabela C.1 - Palavras Sementes para Economia.	101
Tabela C.2 - Palavras Sementes para Educação (parte 01).	102
Tabela C.3 - Palavras Sementes para Educação (parte 02).	103
Tabela C.4 - Palavras Sementes para Empreendedorismo.	103
Tabela C.5 - Palavras Sementes para Energia.	104
Tabela C.6 - Palavras Sementes para Governança.	105
Tabela C.7 - Palavras Sementes para Meio Ambiente.	106

Tabela C.8 - Palavras Sementes para Mobilidade.	106
Tabela C.9 - Palavras Sementes para Saúde.	107
Tabela C.10 -Palavras Sementes para Tecnologia & Inovação.	107
Tabela C.11 -Palavras Sementes para Urbanismo.	108
Tabela D.1 - Experimento I.	110
Tabela D.2 - Experimento II.	111
Tabela D.3 - Experimento III.	112
Tabela D.4 - Experimento IV.	113
Tabela D.5 - Experimento V.	114
Tabela D.6 - Experimento VI.	115
Tabela D.7 - Experimento VII.	116
Tabela D.8 - Experimento VIII.	117
Tabela D.9 - Experimento IX.	118
Tabela D.10 -Experimento X.	119
Tabela E.1 - Termos Representativos para o aspecto Economia. Parte 1.	122
Tabela E.2 - Termos Representativos para o aspecto Economia. Parte 2.	123
Tabela E.3 - Termos Representativos para o aspecto Economia. Parte 3.	124
Tabela E.4 - Termos Representativos para o aspecto Educação. Parte 1.	125
Tabela E.5 - Termos Representativos para o aspecto Educação. Parte 2.	126
Tabela E.6 - Termos Representativos para o aspecto Educação. Parte 3.	127
Tabela E.7 - Termos Representativos para o aspecto Empreendedorismo.	128
Tabela E.8 - Termos Representativos para o aspecto Energia.	128
Tabela E.9 - Termos Representativos para o aspecto Governança. Parte 1.	129
Tabela E.10 -Termos Representativos para o aspecto Governança. Parte 2.	130
Tabela E.11 -Termos Representativos para o aspecto Governança. Parte 3.	131
Tabela E.12 -Termos Representativos para o aspecto Meio Ambiente. Parte 1.	132
Tabela E.13 -Termos Representativos para o aspecto Meio Ambiente. Parte 2.	133
Tabela E.14 -Termos Representativos para o aspecto Mobilidade.	134
Tabela E.15 -Termos Representativos para o aspecto Saúde. Parte 1.	135
Tabela E.16 -Termos Representativos para o aspecto Saúde. Parte 2.	136
Tabela E.17 -Termos Representativos para o aspecto Segurança. Parte 1.	137

Tabela E.18 -Termos Representativos para o aspecto Segurança. Parte 2.	138
Tabela E.19 -Termos Representativos para o aspecto Segurança. Parte 3.	139
Tabela E.20 -Termos Representativos para o aspecto Tecnologia & Inovação.	140
Tabela E.21 -Termos Representativos para o aspecto Urbanismo.	140
 Tabela G.1 - Termos Representativos para o aspecto Economia. Parte 1.	152
Tabela G.2 - Termos Representativos para o aspecto Economia. Parte 2.	153
Tabela G.3 - Termos Representativos para o aspecto Economia. Parte 3.	154
Tabela G.4 - Termos Representativos para o aspecto Educação	154
Tabela G.5 - Termos Representativos para o aspecto Empreendedorismo.	155
Tabela G.6 - Termos Representativos para o aspecto Energia.	155
Tabela G.7 - Termos Representativos para o aspecto Governança. Parte 1.	156
Tabela G.8 - Termos Representativos para o aspecto Governança. Parte 2.	157
Tabela G.9 - Termos Representativos para o aspecto Meio Ambiente. Parte 1.	158
Tabela G.10 -Termos Representativos para o aspecto Meio Ambiente. Parte 2.	159
Tabela G.11 -Termos Representativos para o aspecto Mobilidade.	160
Tabela G.12 -Termos Representativos para o aspecto Saúde.	161
Tabela G.13 -Termos Representativos para o aspecto Segurança.	162
Tabela G.14 -Termos Representativos para o aspecto Urbanismo.	162
 Tabela A.1 - Indicadores de Cidade Inteligente propostos pela <i>Urban Systems</i> . Parte 1.	174
Tabela A.2 - Indicadores de Cidade Inteligente propostos pela <i>Urban Systems</i> . Parte 2.	175

LISTA DE ABREVIATURAS E SIGLAS

API	Application Programming Interface
EI	Extração de Informações
EMIM	Expected Mutual Information Measure
GB	Gigabytes
IDE	Integrated Development Environment
JSON	Java Script Object Notation
KDD	Knowledge Discovery in Databases
KDT	Knowledge Discovery in Text
LSI	Latent Semantic Indexing
PLN	Processamento de Linguagem Natural
SRI	Sistema de Recuperação de Informações
SQL	Structured Query Language
SVM	Support Vector Machine
TIC	Tecnologia da Informação e Comunicação

1 INTRODUÇÃO

Os processos de explorar grandes conjuntos de dados a fim de encontrar relacionamentos inesperados e de resumir os dados de uma forma que eles sejam tanto úteis quanto compreensíveis são algumas das tarefas da mineração de dados, que consiste em um campo interdisciplinar e utiliza técnicas de máquinas de conhecimentos, reconhecimento de padrões, estatísticas, banco de dados e visualização, para conseguir extrair informações de grandes bases de dados (CABENA et al., 1997).

Nos últimos anos, as pesquisas científicas têm voltado o foco para dados armazenados em meio digital sem a preocupação com o rigor da estruturação – os documentos textuais – comumente chamados de “dados não estruturados”. O conteúdo destes documentos, muitas vezes, pode conter informações importantes, “camoufladas” em grandes volumes de textos, que precisam ser explorados em busca de padrões até então desconhecidos, para tomada de decisão e geração de novos conhecimentos (RAMOS; BRÄSCHER, 2009).

No mundo em que vivemos, cada vez mais pessoas têm acesso sem restrições à internet e podem divulgar suas ideias através das suas diferentes mídias. Isto faz com que exista um crescente volume de postagens pela Web, principalmente nas redes sociais. Estas postagens podem tratar-se apenas de publicações e comentários na rede social Facebook¹, tais como notícias, previsão do tempo, enquetes entre outros, ou de artigos sobre as mais diversas áreas.

Diariamente são postadas em redes sociais, publicações sobre assuntos das mais diversas naturezas. Por exemplo, podem estar relacionadas à qualidade da gestão pública de um município, ou até mesmo a um determinado jogo de futebol. Esses posicionamentos podem ser utilizados de forma estratégica por diferentes organizações interessadas em obter informações sobre determinado assunto ou tema.

Publicações em redes sociais contêm informações que, muitas vezes, podem expressar percepções relevantes a serem consideradas para a tomada de decisão, por exemplo, na gestão pública, para o entendimento das manifestações da população em redes sociais. A implementação de inovações em tecnologias da informação e comunicação (TICs) pode conduzir a uma maior eficiência nas atividades relacionadas à prestação dos serviços públicos que envolvem a

¹ <https://www.facebook.com>

gestão de um município, com importantes resultados em favor dos atores que nele estão inseridos (WEISS, 2013).

O modelo de *cidade inteligente* estabelece-se como o monitoramento e integração das condições de operações das infraestruturas críticas da cidade, atuando de forma preventiva para a continuidade de atividades essenciais, melhorando as condições de serviços e a qualidade de vida dos cidadãos (HALL et al., 2000).

Uma cidade inteligente é sinônimo de um município sustentável apto a produzir, consumir e distribuir um grande número de informações, geralmente em tempo real. Ela utiliza a tecnologia para prestar, de forma mais eficiente, os serviços urbanos, melhorar a qualidade de vida das pessoas e transformar a relação entre entidades locais, empresas e cidadãos proporcionando uma nova forma de viver na cidade. Ademais, atende às demandas de forma integrada e o mais rápido possível. Este modelo de processamento inteligente serve como referência às tomadas de decisões de empresas, governos e cidadãos, com o intuito de tornar as atividades urbanas mais eficientes e sustentáveis nas esferas econômica, social, ecológica e política.

Atualmente, diversos indicadores estão sendo propostos para avaliar o grau de inteligência de uma cidade (SYSTEM, 2017). Estas métricas relacionam-se principalmente a dados quantitativos associados às questões de infraestrutura e suas tecnologias. Estes indicadores estão sendo utilizados pelos gestores públicos, bem como permitem o estabelecimento de métricas de comparação entre diferentes municípios.

Entretanto, informações quantitativas sobre o grau de infraestrutura e suas tecnologias pouco podem expressar a verdadeira apropriação e uso destes pelos seus cidadãos. Por exemplo, qual a importância em dados quantitativos de linhas de ônibus monitorados se a população está descontente com este serviço?

Neste contexto, percebe-se que existe uma lacuna escondida entre os indicadores atualmente utilizados no estabelecimento do grau de inteligência de uma cidade e no efetivo nível de satisfação do seu cidadão, o que em última análise, equivaleria ao grau de otimização da sua inteligência. Essa lacuna poderia ser preenchida pelo entendimento das diferentes percepções de seus cidadãos com relação aos diferentes aspectos constituintes da sua vivência nesta cidade. Este entendimento poderia trazer uma riqueza de representação atualmente não capturada pelos indicadores de cidades inteligentes disponíveis. Onde buscar estas percepções? Como representá-las? Em tempos de redes sociais, estes locais tornam-se espaços onde os indi-

víduos manifestam comentários relacionados a experiências do seu cotidiano. A escala destas manifestações cresce a cada dia, transformando-se em uma fonte de informações, com muitas potencialidades a ser explorada. Diante deste cenário, a proposta deste trabalho é estudar e propor técnicas computacionais capazes de avaliar a percepção dos cidadãos a respeito da sua cidade, a partir do tratamento e análise de postagens em redes sociais. Com base na literatura de cidades inteligentes e seus indicadores, busca-se pré-procesar e minerar grandes quantidades de postagens no Facebook com o intuito de fornecer termos representativos a serem tratados pelos gestores públicos. A proposta será aplicada em um estudo de caso envolvendo postagens em diferentes grupos e páginas do Facebook, relacionadas às informações da cidade do Rio Grande / RS. Busca-se obter termos representativos, a serem tratados por seus gestores relativos aos seguintes aspectos: Economia, Educação, Empreendedorismo, Energia, Governança, Meio Ambiente, Mobilidade, Saúde, Segurança, Tecnologia & Inovação e Urbanismo. Estes aspectos são considerados indicadores importantes, contabilizados em muitos dos rankings nacionais e internacionais de aferição da qualidade de vida urbana das cidades. Busca-se assim propor o uso de técnicas de classificação baseadas em termos representativos para que a partir destes possam ser extraídos mapas conceituais sobre os principais indicadores que compoem o nível de inteligência de um município. Com este estudo pretende-se analisar dados gerados pela população em meios as redes sociais sobre diferentes aspectos e gerar resultados que poderão ser utilizados por agentes públicos para auxiliar à tomada de decisão. O estudo avaliará conceitos relacionados à mineração de textos, por meio de um modelo computacional de extração, pré-processamento, mineração e pós-processamento dos dados textuais. Ademais, essa investigação assume o desafio tecnológico de realizar a classificação em textos de publicações fracamente estruturadas, considerando que os dados textuais contêm pouca ou quase nenhuma estrutura.

1.1 MOTIVAÇÃO E JUSTIFICATIVA

O desenvolvimento desse trabalho tem como motivação a oportunidade de realizar uma pesquisa com enfoque na Computação Social, a qual busca aplicar técnicas de mineração de dados com objetivo de gerar resultados, a partir de dados expressos pela população na rede social Facebook, sobre aspectos e seus indicadores sociais, no contexto de cidades inteligentes. Essas informações coletadas, podem representar uma fonte de dados para o apoio à gestão

pública na tomada de decisão.

A justificativa desta pesquisa se insere nos seguintes pontos:

- Quantidade de pessoas que usam redes sociais, conforme apresenta a Figura 1.1;

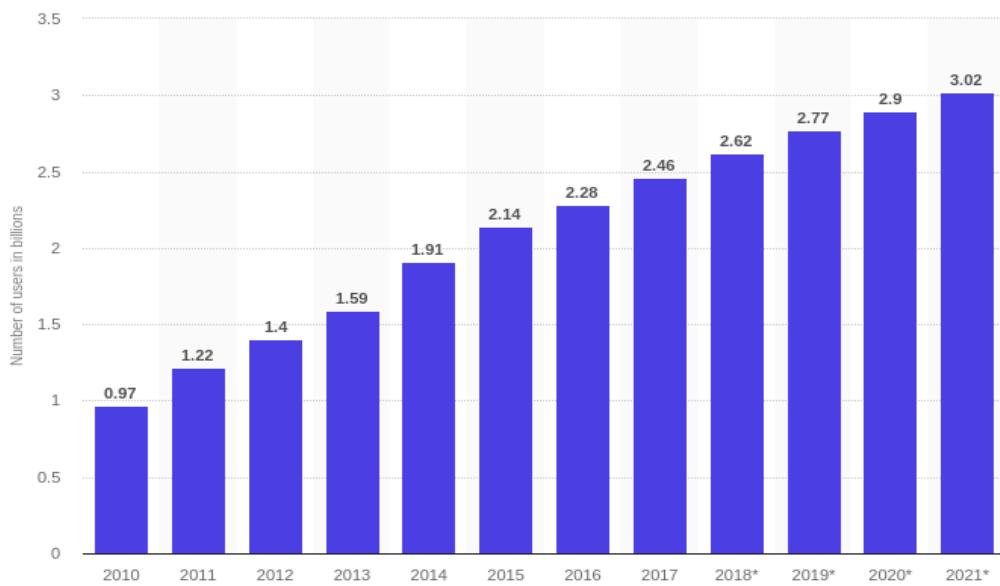


Figura 1.1: Número de usuários de mídias sociais em todo o mundo no período de 2010 a 2021 (em bilhões). Fonte: (STATISTA, 2017a).

- Quantidade de pessoas que usam Facebook no Brasil, de acordo com os dados apresentados na Figura 1.2;

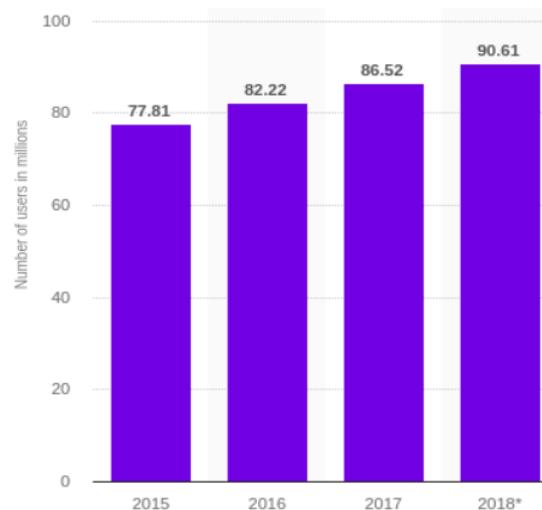


Figura 1.2: Número de usuários do Facebook no Brasil de 2015 a 2018 (em milhões). Fonte: (STATISTA, 2017b).

- Popularidade do Facebook no Brasil, sendo a maior rede social utilizada. Estes dados são apresentados na Figura 1.3;

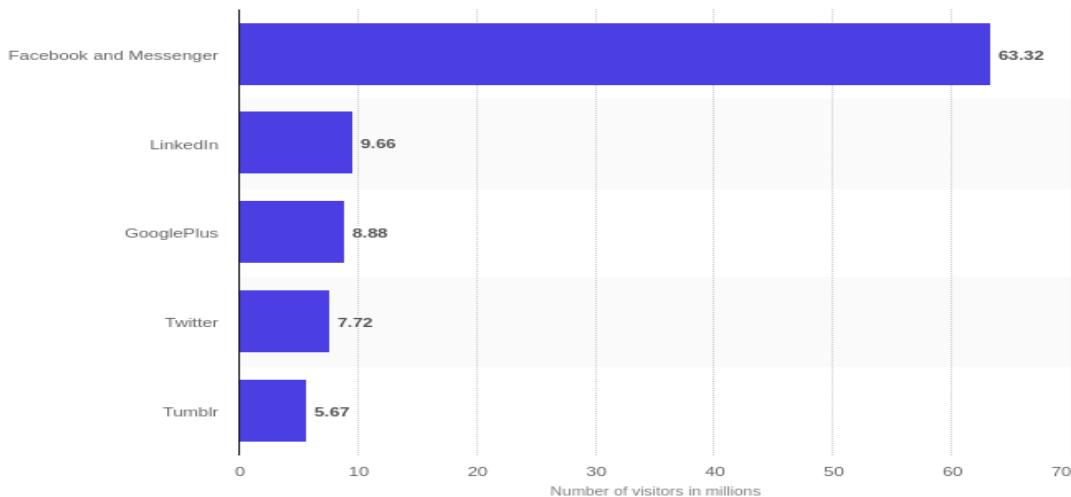


Figura 1.3: Redes sociais mais populares no Brasil a partir de julho de 2015, com base no número de visitantes únicos (em milhões). Fonte: (STATISTA, 2017c).

- Ausência de limite de caracteres em publicações do Facebook, que ao contrário do Twitter², só permite 280 caracteres.

1.2 OBJETIVOS

A seguir serão apresentados os objetivos deste trabalho.

1.2.1 Objetivo Geral

Esta pesquisa tem como objetivo estudar e propor uma metodologia baseada em classificação de textos a ser aplicada em postagens de redes sociais, com o intuito de obter indícios sobre a percepção coletiva da população a partir de termos representativos relacionados à indicadores de qualidade de vida no âmbito de cidades inteligentes. A proposta será validada em um estudo de caso associado a dados textuais extraídos de páginas e grupos temáticos do Facebook, relacionados às informações da cidade do Rio Grande / RS.

² <https://twitter.com>

1.2.2 Objetivos Específicos

- Realizar estudo bibliográfico referente a indicadores para cidades inteligentes e mineração de textos.
- Propor um modelo para obtenção de termos representativos relacionados a aspectos e indicadores de cidades inteligentes.
- Aplicar o modelo a um estudo de caso envolvendo 40 páginas e grupos temáticos e 53579 postagens de publicações relacionadas às informações da cidade do Rio Grande / RS, disponibilizando um *dataset* anotado com 1496 mil postagens para teste e validação do método.

1.3 ORGANIZAÇÃO

Este trabalho está organizado da seguinte maneira. O Capítulo 2 apresenta o referencial teórico abordando inicialmente indicadores para cidades inteligentes, seguido de conceitos, definições e técnicas de mineração de dados e mineração de textos. No capítulo 3 os trabalhos relacionados são apresentados. Já o capítulo 4 apresenta a proposta metodológica para classificação. A metodologia é aplicada e validada no estudo de caso referente à mineração de postagens do Facebook para inferência da inteligência do território, o qual considera o município do Rio Grande / RS, apresentada no Capítulo 5. Por fim, no Capítulo 6 considerações finais são apresentadas, bem como trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 CIDADES INTELIGENTES

De acordo com Caragliu, Bo e Nijkamp (2011), uma cidade pode ser definida como inteligente quando há um investimento humano, de capital e na infra-estrutura de tecnologia da informação e comunicação. Para Komninos (2013), consiste em um ambiente que engloba um espaço digital, através das TIC's, associado ao processamento de informações, a transferência de conhecimento e a instrumentos vinculados a tecnologia.

Existem várias definições na literatura sobre cidades inteligentes. De acordo com Giffinger e Gudrun (2010), são definidas pela eficiente visão de futuro em várias vertentes: economia, pessoas, governança, mobilidade, meio ambiente e qualidade de vida.

Segundo Hall et al. (2000), cidades inteligentes são aquelas que monitoram e integram as condições de operações de todas as infraestruturas críticas da cidade, atuando de forma preventiva para a continuidade de suas atividades fundamentais. Já para Harrison e Donnelly (2011), são cidades que fazem uso sistemático das TICs para promover a eficiência no planejamento, execução e manutenção dos serviços e infraestruturas urbanas.

A próxima seção deste capítulo, no contexto de cidades inteligentes, apresenta os eixos utilizados em rankings nacionais e internacionais de aferição da performance urbana das cidades, que neste trabalho são denominados como aspectos.

2.2 INDICADORES PARA CIDADES INTELIGENTES

Para a delimitação da unidade-caso, visando o contexto de Cidade Inteligente, nessa pesquisa, através da mineração de textos, propõem-se a apresentar informações que poderão ser utilizados por agentes públicos para auxiliar a tomada de decisão. Estas informações referem-se a aspectos importantes, contabilizados em muitos dos rankings nacionais e internacionais de aferição da performance urbana das cidades; visando o entendimento da população sobre os mesmos em meio as redes sociais como o Facebook.

Dentre os índices mais divulgados, destaca-se o *European Smart Cities* (Europeu),

metodologia que descreve a cidade inteligente através de seis (6) características (economia inteligente, pessoas inteligentes, governança inteligente, mobilidade inteligente, meio ambiente inteligente, vida inteligente), 31 fatores e 74 indicadores (KHATOUN; ZEADALLY, 2016). No Brasil, a *Urban Systems*, apresenta o *Ranking Connected Smart Cities* (SYSTEM, 2017), que considera o desenvolvimento de 11 eixos ou aspectos, sendo eles: Mobilidade, Urbanismo, Meio Ambiente, Energia, Tecnologia e Inovação, Economia, Educação, Saúde, Segurança, Empreendedorismo e Governança. Além disso, são utilizados 70 indicadores, os quais foram pensados e estudados para atender a um dos 11 eixos da pesquisa. No entanto, alguns indicadores estão relacionados a mais de um eixo. Os indicadores desta pesquisa estão ilustrados através das Tabelas A.1 e A.2, localizados no anexo deste documento.

Outras iniciativas vêm sendo consideradas, como por exemplo, *IESE: Cities in motion index 2016* (BERRONE et al., 2016) e no Brasil, o projeto “Brasil 2030: Cidades inteligentes e humanas”, coordenado pela Rede Brasileira de Cidades Inteligentes e Humanas, uma parceria entre a Frente Nacional de Prefeitos e diversas instituições. Entretanto, tais ações ainda configuram-se como propostas a serem validadas pela aplicação e análise no território nacional. A Tabela 2.1 apresenta uma análise comparativa dos *Rankings* de Inteligência Urbana.

Dentre as empresas analisadas, durante 20 anos, a *Urban Systems*, com projetos em mais de 800 cidades, vem acumulando conhecimento e competência em diversos segmentos da economia, que a tornaram referência em inteligência de mercado e consultoria de negócios em todo Brasil (SYSTEMS, 2018). Diante deste cenário, os eixos apresentados por esta Instituição serão considerados como aspectos para o processo de mineração de textos neste estudo, devido a consolidação desta empresa em pesquisas realizadas no território brasileiro. Ademais, nesta pesquisa considera-se apenas uma entidade: a cidade do Rio Grande/RS.

Tabela 2.1: Análise comparativa dos Rankings de Inteligência Urbana

		RANKING	
		EUROPEAN SMART CITIES	BRASIL 2030
		RANKING CONNECTED SMART CITIES	
ITENS			
Metodologia	- 2012: Aplicação em 20 municípios - 2016: Atualização para análise de 36 municípios da rede.	Ainda não aplicado	Foram 700 municípios analisados em todo o país.
Eixos	Economia inteligente, pessoas inteligentes, governança inteligente, mobilidade inteligente, meio ambiente inteligente, vida inteligente.	Governança, arquitetura, urbanismo e antropologia, tecnologia e segurança.	Mobilidade, urbanismo, meio ambiente, energia tecnologia e inovação, economia, educação, saúde, segurança, empreendedorismo e governança.
Fatores	31 fatores	-	-
Indicadores	74 indicadores	205 indicadores	70 indicadores
Instituição	- Centro de Ciência Regional da Universidade de Tecnologia de Vienna; - Departamento de Geografia da Universidade de Ljubljana; - Instituto de Pesquisa para Habitação, Urbanismo e Estudos de Mobilidade da Universidade de Tecnologia Delft; - RENER – Rede Portuguesa de Cidades Inteligentes	- Rede Brasileira de Cidades Inteligentes & Humanas (RBCS)	Urban Systems
Ano	2012	2017	2016
Atualização	2016	2017	2017
Tipo	PRIVADO – INTELI	PÚBLICO	PRI VADO
Abrangência	COMUNIDADE EUROPÉIA	Brasil	Brasil
Objetivos	Aborda explicitamente cidades de médio porte na Europa, levando em consideração suas perspectivas e desafios de desenvolvimento.	Criar o Índice Brasileiro de Cidades Inteligentes e Humanas	Avaliar o potencial de desenvolvimento das cidades brasileiras.

2.3 MINERAÇÃO DE DADOS

A mineração de dados é o processo de obter e interpretar informações, através de modelos preditivos, aplicados a grandes bases de dados. Consiste em uma das etapas do processo de Descoberta de Conhecimento em Bases de Dados (*Knowledge Discovery in Databases – KDD*), composta por um conjunto de ferramentas e técnicas computacionais, que implementam algoritmos de aprendizagem ou classificação.

De acordo com Han, Pei e Kamber (2011), a mineração de dados é um campo interdisciplinar que utiliza conceitos de bancos de dados, estatística, aprendizagem de máquina, inteligência artificial, reconhecimento de padrões, sistemas de armazenamento de dados, recuperação da informação, visualização, algoritmos de computação de alto desempenho e diversos outros domínios de aplicação, conforme ilustrado na Figura 2.1.

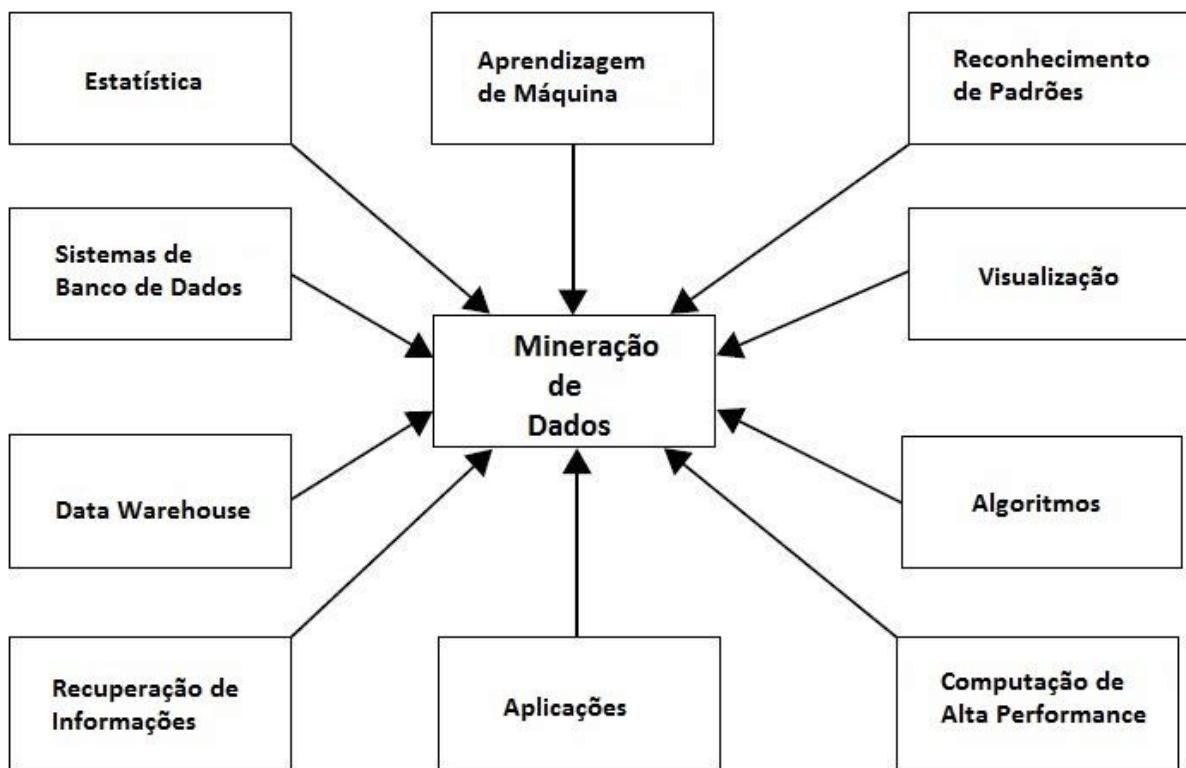


Figura 2.1: Mineração de dados e técnicas de diversas áreas relacionadas. Fonte adaptada de Han, Pei e Kamber (2011).

O processo de mineração de dados pode ser dividido em: tarefas de previsão e tarefas descritivas. As tarefas de previsão têm como objetivo prever o valor de um determinado atributo com base nos valores de outros atributos. Já nas tarefas descritivas, o objetivo é derivar padrões

(TAN; STEINBACH; KUMAR, 2005). O autor destaca que as tarefas descritivas frequentemente necessitam de técnicas de pós-processamento para avaliar os resultados obtidos.

As principais etapas da mineração de dados são a análise exploratória, a descoberta de padrões frequentes, o agrupamento e a classificação de dados (MADEIRA, 2015).

Segundo Madeira (2015), a análise exploratória dos dados consiste em processar os dados numéricos e categóricos, de modo a extrair características chave da amostra por meio de estatísticas. Já a descoberta de padrões frequentes está associada à extração de padrões informativos a partir de dados complexos e em larga escala. Na clusterização (*clustering*) ou agrupamento, os dados são separados em grupos denominados clusters. Por fim, a classificação é a etapa que consiste em predizer a categoria ou classe de um registro não classificado. É importante destacar que para construir um classificador, é necessário um conjunto de registros classificados corretamente. Esse conjunto é chamado conjunto de treinamento. Após o aprendizado através dos registros do conjunto de treinamento, o classificador é capaz de prever automaticamente a classe de qualquer novo registro. Portanto, diferente do agrupamento, a classificação é um processo de aprendizado supervisionado.

Basicamente, a mineração de dados é uma das etapas de um processo de Descoberta de Conhecimento (KD), dividido em duas abordagens: a Descoberta de Conhecimento em Bases de Dados Estruturados e a Descoberta de Conhecimento em Dados não Estruturados (*Knowledge Discovery from Text*), ilustradas através da Figura 2.2 (MORAIS; AMBRÓSIO, 2007).

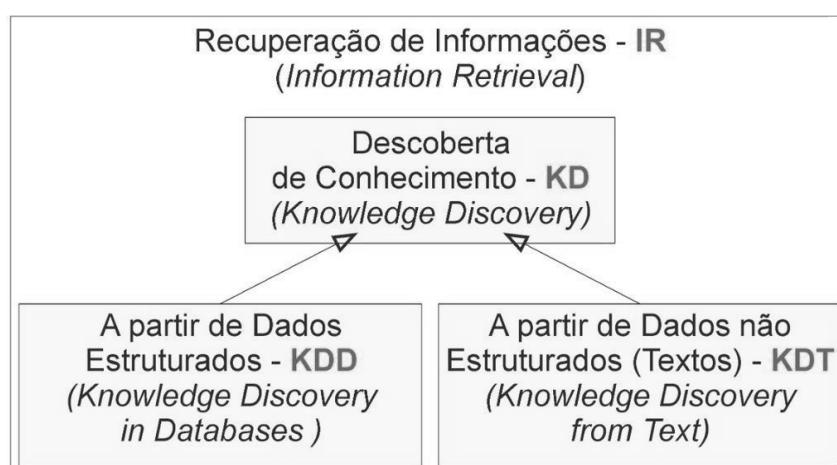


Figura 2.2: Abordagens utilizadas no processo de Descoberta de Conhecimento. Fonte adaptada de Morais e Ambrósio (2007).

De acordo com Wives (2002), descobrir conhecimento significa identificar informa-

ções relevantes, processá-las e agregá-las ao conhecimento prévio do indivíduo, alterando o estado de conhecimento atual, com o intuito de resolver determinada situação ou problema.

De maneira geral, organizações de diversas naturezas, armazenam em diferentes formatos e estrutura, grande volume de dados e informações, que são constantemente demandados. Como os volumes de dados crescem dramaticamente, este tipo de análise, manual, torna-se impraticável em vários domínios (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996a).

A análise de grandes quantidades de dados pelo homem é inviável sem o auxílio de ferramentas computacionais apropriadas. Portanto, torna-se imprescindível o desenvolvimento e a utilização de ferramentas que auxiliem o homem, de forma automática e inteligente, nas tarefas de analisar, interpretar e relacionar dados, principalmente quando essas informações podem ser representativas para o desenvolvimento e a seleção de estratégias de tomada de decisão em seu contexto de aplicação (PASSOS; GOLDSCHMIDT, 2005).

Segundo Fayyad, Piatetsky-Shapiro e Smyth (1996a), historicamente a Descoberta de Conhecimento em Banco de Dados (Knowledge Discovery in Databases - KDD) foi formalizada em 1989, no primeiro workshop de KDD. Estabelece-se como processo organizado a partir de duas classes de objetivos: a Verificação e a Descoberta. Na verificação ocorre a confirmação ou não de uma hipótese formulada pelo pesquisador, enquanto na descoberta ocorre a busca por novos conhecimentos, de forma automática.

O processo de KDD é composto por uma série de etapas que são apresentadas na Figura 2.3 reproduzida de Passos e Goldschmidt (2005) e descritas abaixo:

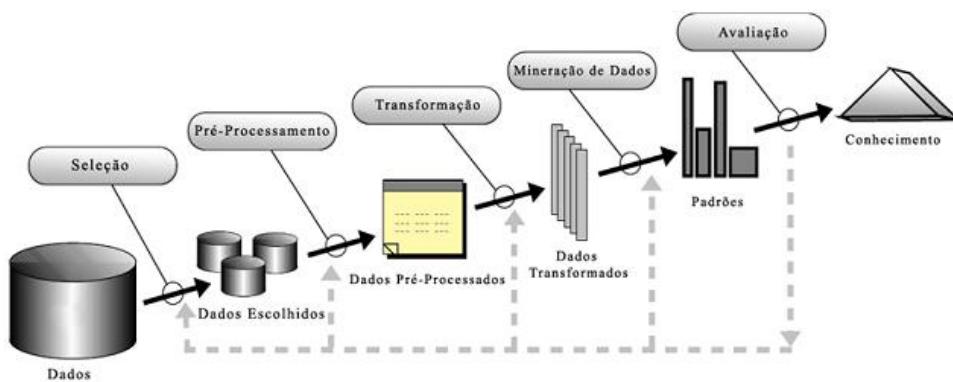


Figura 2.3: Visão geral das etapas do processo de Descoberta de Conhecimento em Base de Dados - KDD. Fonte adaptada de Passos e Goldschmidt (2005).

- Seleção de dados: etapa na qual é definido um conjunto de dados-alvo para serem analisa-

dos. Estes dados são selecionados de acordo com os objetivos do processo de descoberta de conhecimento que será realizado.

- Pré-processamento: essa fase tem como objetivo tratar os dados previamente selecionados, realizando a limpeza e remoção de ruídos dos mesmos.
- Transformação: nessa etapa, os dados são transformados de acordo com a tarefa de mineração de dados. Fazem parte dessa fase a redução de dimensionalidade do conjunto de dados (quantidade de atributos), a redução de registros existentes, entre outros.
- Mineração de dados: consiste no processo de exploração de grandes quantidades de dados com o objetivo de encontrar anomalias, padrões e correlações para suportar a tomada de decisões e proporcionar vantagens estratégicas. Nesta etapa, conforme o objetivo do processo, é escolhida uma tarefa de mineração. O algoritmo é escolhido e parametrizado de acordo com as características do conjunto de dados-alvo.
- Interpretação e avaliação dos resultados: os resultados obtidos são interpretados e avaliados. Medições estatísticas ou métodos de teste de hipóteses podem ser aplicados com o objetivo de eliminar resultados não legítimos.

Fayyad, Piatetsky-Shapiro e Smyth (1996b) destacam que a mineração de dados é uma das etapas do processo de descoberta de conhecimento em bases de dados estruturados. Ela surge quando uma organização, de qualquer natureza, identifica a necessidade de analisar e utilizar um volume de dados disponíveis obtendo informações até então implícitas ou não triviais. Entretanto, é importante observar que nem toda informação disponível em grande volume de dados, é concebida de forma estruturada. Os dados não estruturados necessitam de mecanismos computacionais diferentes dos estruturados, para que possam ser coletados, armazenados, manipulados e consultados.

A análise de dados armazenados em formato não-estruturado pode ser considerada uma atividade mais complexa, se comparada à análise de dados estruturados, justamente pelo fato desses dados exigirem um processo de pré-definição de estrutura. Logo, são necessárias técnicas e ferramentas específicas para tratamento deste tipo de dados. Esse conjunto de procedimentos específicos fazem parte da área de Recuperação de Informações, mais especificamente na Descoberta de Conhecimento em Textos - KDT (PALAZZO et al.,); (WIVES, 2005). Entre-

tanto, vale citar que dados não-estruturados podem estar representados em formatos de imagem, audio ou vídeo.

Neste trabalho, os dados não-estruturados que serão aplicados em um estudo de caso, são textos em formato livre e fracamente estruturados (que contêm pouca ou nenhuma estrutura). Estes textos referem-se às publicações na rede social Facebook que foram postadas pela população em páginas e grupos temáticos relacionados às informações da cidade do Rio Grande / RS. Geralmente, esses textos contêm erros gramaticais e por este motivo são fracamente estruturados. Além disso, este trabalho tem foco na classificação de dados, considerando que pretende-se classificar as publicações ou documentos textuais.

Na próxima seção, serão apresentados conceitos e definições da mineração de textos, técnica semelhante a mineração de dados, porém aplicada sobre dados textuais, geralmente não-estruturados, o qual é um dos objetivos de estudo deste trabalho.

2.4 MINERAÇÃO DE TEXTOS

A mineração de textos, também conhecida como Descoberta de Conhecimento em Textos - KDT, refere-se ao processo de extrair informações úteis a partir de uma grande base de dados textuais não estruturados. Ela é considerada como uma evolução dentro da área de Recuperação de Informações, podendo ser vista como uma extensão da mineração de dados, onde trata-se de um campo novo e interdisciplinar, que inclui conhecimentos de áreas como Informática, Estatística, Linguística e Ciência Cognitiva.

O objetivo da mineração de textos é identificar padrões em grandes bases de dados textuais não estruturados escritos em linguagem natural, de modo a gerar conhecimento a partir da extração de informações úteis e não triviais, que muitas vezes não são evidentes nos textos e que não poderiam ser obtidos através de métodos tradicionais de pesquisa.

Existem várias definições na literatura relacionadas à mineração de textos. De acordo com Lopes (2004), a mineração de textos refere-se ao processo de extração de padrões interessantes e não triviais, ou conhecimento a partir de documentos em textos não-estruturados.

Moura (2004) descreve a mineração de textos como sendo uma área de pesquisa tecnológica cujo objetivo é a busca por padrões, tendências e regularidades em textos escritos em linguagem natural.

Já Salton e McGill (1986) e Wives (2002) afirmam que a mineração de textos é um processo de descoberta de conhecimento que se utiliza de ferramentas computacionais e algoritmos específicos, de modo a processar e identificar textos, frases e palavras. Wives (2002) ainda afirma que a mineração de textos pode ser entendida como a aplicação de técnicas de KDD sobre dados extraídos de textos. Entretanto, KDT não inclui somente a aplicação das técnicas tradicionais de KDD, mas também técnicas recentes que possam ser aplicadas no sentido de encontrar conhecimento em qualquer tipo de texto.

As principais contribuições desta área estão relacionadas à busca de informações específicas em documentos, à análise qualitativa e quantitativa de grandes volumes de textos, e à melhor compreensão de textos disponíveis em documentos. Na prática, a mineração de textos define um processo inovador que auxilia na descoberta de conhecimento a partir de documentos textuais, que pode ser utilizado em diversas áreas (MORAIS; AMBRÓSIO, 2007).

Conforme demonstra os estudos de Morais e Ambrósio (2007), um processo de mineração de textos pode ser dividido nas seguintes etapas: seleção de documentos, definição do tipo de abordagem dos dados (análise semântica ou estatística), preparação dos dados, indexação e normalização, cálculo da relevância dos termos, seleção dos termos e pós-processamento (análise de resultados). As etapas são ilustradas através da Figura 2.4.

2.4.1 Seleção dos Textos e Tipos de Abordagens de Dados

A identificação do problema é a etapa onde são selecionados os textos e definida a abordagem de análise a ser utilizada. De acordo com os estudos de Morais e Ambrósio (2007) e Junior (2007), existem dois tipos de análises em um processo de mineração de textos, a análise semântica e a análise estatística.

A análise semântica avalia a sequência de termos no contexto da frase, com o objetivo de identificar a função correta de cada termo, por meio de técnicas de Processamento de Linguagem Natural (*Natural Language Processing - PLN*) (CORDEIRO et al., 2005).

Já a análise estatística busca extrair informações com base na frequência dos termos encontrados nos textos, sendo a importância de um termo o número de vezes que ele aparece em um texto. Este tipo de análise envolve aprendizado estatístico a partir de dados, e pode incluir algumas etapas, como codificação dos dados, estimativa dos dados e modelos de representação

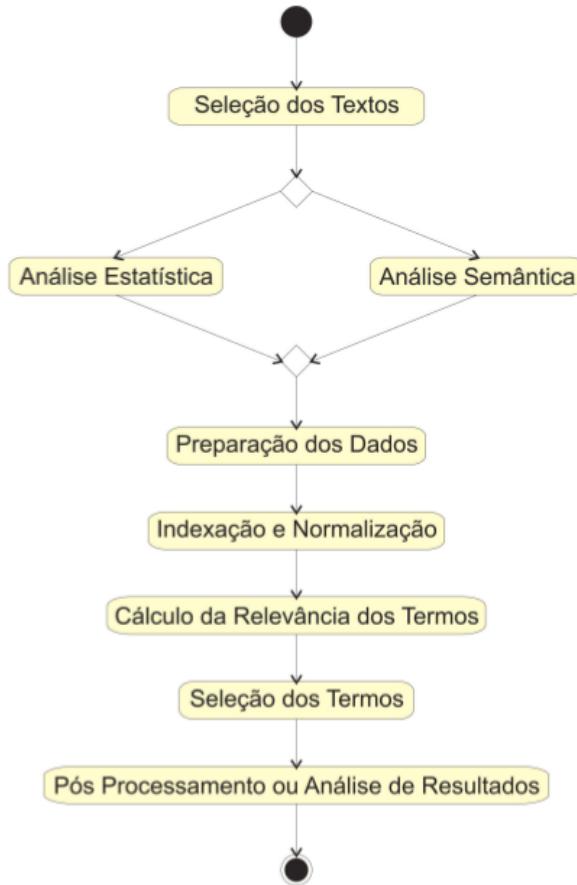


Figura 2.4: Etapas do processo de mineração de textos. Fonte: Morais e Ambrósio (2007).

de documentos (MORAIS; AMBRÓSIO, 2007).

Este trabalho considera a análise estatística para o processo de mineração de textos, a qual consiste em classificar os documentos em aspectos.

2.4.2 Preparação dos Dados

Na etapa de preparação dos dados, os documentos textuais são analisados de modo a extrair dos dados não estruturados, uma representação estruturada. Normalmente, os documentos são representados por vetores de atributos, sendo atributos um associativo entre cada palavra existente no texto e o valor de seu peso, o qual é calculado através da frequência que a palavra ocorre no documento de texto que está sendo analisado. Dessa maneira, a base de documentos é armazenada em estrutura de matriz ou grafo e cada documento pode ser representado em um plano, possibilitando cálculos de funções de similaridade e análise de relevância sobre este conjunto de documentos, o que permite retornar a um usuário resultados que correspondam à sua

necessidade de busca. De acordo com os estudos de Morais e Ambrósio (2007) e Junior (2007), existem vários modelos para o cálculo da similaridade, muitos deles são oriundos da recuperação de informação, como o booleano, o espaço-vetorial, o probabilístico, o difuso (*fuzzy*), o da busca direta, o de aglomerados (*clusters*), o lógico e, o contextual ou conceitual.

Esta pesquisa utiliza o modelo booleano e vetorial para a preparação dos dados. No entanto, não propõe-se a realizar cálculos de similaridade entre termos e documentos.

2.4.3 Indexação e Normalização

Na etapa de indexação e normalização, o objetivo é facilitar a identificação de similaridade de significado entre suas palavras, considerando as variações morfológicas e problemas de sinonímia (EBECKEN et al., 2003). O resultado é a geração de um índice, construído através de um processo de indexação. Nesse sentido, o processo de indexação consiste em analisar o conteúdo de um documento e atribuir a ele um índice, ou seja, classificá-lo em determinado assunto ou tema. Dessa forma, os documentos ficam organizados através de índices, facilitando o acesso rápido à informação.

Segundo Morais e Ambrósio (2007), um documento pode ser indexado por diferentes termos correspondentes ao vocabulário utilizado na área. Nesse caso, geralmente, há um conjunto de termos predefinidos e específicos para cada assunto da área em questão.

De maneira geral, um processo de indexação pode ser realizado manualmente ou de forma automática, sendo comparado as etapas de classificação e agrupamento da mineração de dados. Quanto realizado manualmente, um especialista no assunto fica encarregado de analisar o conteúdo dos documentos e classificá-los em determinado(s) índice(s), de modo a gerar um treinamento para futuras classificações, sendo estas de maneira automática. Ou seja, trata-se de um processo de aprendizagem de máquina supervisionado. Já na indexação automática, os dados são separados por grupos denominados *clusters*, através de técnicas de agrupamento, processo de aprendizagem de máquina não supervisionado, que analisa a similaridade entre os documentos. O resultado desse agrupamento, pode ser considerado uma base de treinamento para o processo de classificação automática.

De acordo com Morais e Ambrósio (2007), em mineração de textos, a indexação é um processo automático, sendo dividido em três etapas, as quais são descritas:

- Identificação de termos: o objetivo principal desta etapa é a identificação dos termos contidos no texto, sejam eles simples ou compostos. Na identificação de termos simples, geralmente são eliminados símbolos e caracteres de controle, além de verificar a ocorrência de erros ortográficos através de um dicionário de termos. Pode-se utilizar nesta etapa um dicionário de sinônimos para normalização do vocabulário. Além disso, os caracteres podem ser todos convertidos para maiúsculo ou minúsculo; múltiplos espaços e tabulações podem ser substituídos por espaços simples; números e datas podem ser padronizados; hífens podem ser eliminados (MORAIS; AMBRÓSIO, 2007). Na identificação de termos compostos, também conhecida como *Word-phrase formation*, busca-se identificar expressões compostas de dois ou mais termos. Esta identificação pode ser feita de duas formas. A primeira envolve a identificação de termos que co-ocorrem com muita frequência em uma coleção de documentos. A segunda consiste na utilização de um dicionário de expressões (WIVES, 2002);
- Remoção de *stop words*: esta etapa consiste na eliminação de *stop words* ou palavras de parada, as quais são palavras comuns encontradas em grande quantidade de textos e que não carregam significado em si próprias, como por exemplo artigos de ligação, preposições, pronomes e advérbios. Ademais, estas palavras dificilmente são utilizadas em uma consulta. Morais e Ambrósio (2007) destaca que além das *stop words*, existem também palavras cuja frequência na coleção de documentos é muito alta. Palavras que aparecem em praticamente todos os documentos de uma coleção não são capazes de discriminar documentos e também não devem fazer parte do índice;
- Normalização morfológica (*Stemming*): Segundo Wives (2002), em alguns casos é interessante eliminar as variações morfológicas de uma palavra. Elas são eliminadas através da identificação do radical de uma palavra. Os prefixos e os sufixos são retirados e os radicais resultantes são adicionados ao índice. Essa técnica de identificação de radicais trata-se de uma normalização linguística denominada lematização ou *stemming*. Além disso, características de gênero, número e grau das palavras são eliminados, sendo várias palavras reduzidas a um único termo, o que pode reduzir o tamanho de um índice. Entretanto, de acordo com Morais e Ambrósio (2007), a aplicação de técnicas de *stemming* ocasionam uma diminuição na precisão das buscas, já que o usuário não consegue mais procurar por uma palavra específica.

Neste trabalho, a classificação dos documentos consiste em um processo de aprendizagem de máquina supervisionado, onde os documentos são classificados em determinado(s) aspecto(s), de modo a gerar um conjutno de treinamento para futuras classificações, sendo estas de maneira automática. Entre as etapas da indexação, neste estudo aplicou-se a identificação de termos simples, de modo a identificar e remover símbolos e caracteres de controle, e também a remoção de stop words. A etapa de normalização morfológica foi considerada desnecessária, uma vez que ocasiona uma diminuição na precisão das buscas. Além disso, diferente do funcionamento no idioma inglês, a normalização morfológica não funciona bem para o idioma português, devido as regras gramaticais e o fato de que palavras com o mesmo radical podem ter significados diferentes.

De acordo com Morais e Ambrósio (2007), após as etapas de indexação e normalização, o processo de mineração de textos é complementado pelo cálculo de relevância e seleção dos termos, as quais são apresentadas nas próximas seções deste trabalho.

2.4.4 Cálculo de Relevância dos Termos

Para o cálculo de relevância dos termos, Morais e Ambrósio (2007) afirmam que nem todas as palavras em um documento de texto possuem a mesma importância. A relevância de uma palavra em relação ao texto em que está inserido pode basear-se na quantidade ocorrências ou na análise estrutural do documento. As análises baseadas em frequência ou quantidade de ocorrências costumam ser as mais utilizadas. O grau de relacionamento de uma palavra com um texto é denominado peso. Segundo Morais e Ambrósio (2007), existem várias fórmulas para cálculo do peso. As mais comuns são baseadas em cálculos simples de frequência: frequência absoluta, frequência relativa e frequência inversa de documentos.

O cálculo de frequência absoluta, também conhecido como frequência do termo ou *term frequency* (TF), considera a quantidade de vezes que o termo aparece no documento.

Já a frequência relativa não considera apenas a quantidade de vezes que a palavra aparece em determinado documento, mas também o tamanho do documento é levado em consideração. A frequência relativa (F_{rel}) de uma palavra x em um documento qualquer é calculada dividindo-se sua frequência absoluta (F_{abs}) pelo número total de palavras no mesmo documento (N):

$$F_{rel}(x) = \frac{F_{abs}(x)}{N}$$

Por fim, a frequência inversa de documentos (*inverse document frequency* - IDF), na qual palavras que aparecem em poucos documentos são mais importantes do que as palavras que aparecem em muitos documentos. A fórmula de frequência relativa considera apenas a quantidade de documentos em que um termo aparece. Com base na informação da frequência absoluta e da frequência de documentos é possível calcular a frequência inversa de documentos. A fórmula mais comum utilizada para cálculo do peso de um termo utilizando a frequência inversa é:

$$Peso_{td} = \frac{Freq_{td}}{DocFreq_{td}}, \text{ onde:}$$

$Peso_{td}$: é o grau de relação entre o termo t e o documento d ;

$Freq_{td}$: número de vezes que o termo t ocorre no documento d ;

$DocFreq_{td}$: número de documentos que o termo t ocorre.

Segundo Wives (2002), não existe estudo que indique a superioridade de uma técnica sobre outra de forma significativa. Porém, algumas são mais adequadas do que outras para certas aplicações ou modelos conceituais.

Neste estudo, para o cálculo de relevância dos termos, considera-se apenas a frequência absoluta das palavras nos documentos.

2.4.5 Seleção dos Termos

Esta etapa consiste em selecionar os termos representativos de um texto ou de um corpus de treinamento, com base em seu peso ou frequência que ocorre em cada documento de texto analisado. Geralmente, são desprezados termos que possuam peso inferior a valores definidos por um especialista no processo de mineração de textos ou a seleção ocorre em quantidade determinada de n termos relevantes.

Segundo Morais e Ambrósio (2007), as principais técnicas de seleção de termos são: filtragem baseada no peso do termo, seleção baseada no peso do termo, seleção por análise de

coocorrência, seleção por *Latent Semantic Indexing - LSI* e seleção por análise de linguagem natural.

2.4.6 Pós-Processamento ou Análise de Resultados

Nesta etapa são aplicadas técnicas para analisar os resultados do classificador, particularmente de um processo de mineração de textos, com o intuito de verificar a eficiência da aplicação dos algoritmos aplicados na etapa anterior. Na prática, são aplicadas métricas para avaliação dos resultados, com o objetivo de observar a eficiência do mecanismo. As métricas podem informar quantos e quais documentos são relevantes para uma consulta. A eficiência dessa forma pode ser medida e avaliada de acordo com a sua capacidade de recuperar o máximo possível de documentos relevantes e ignorar o maior número de documentos irrelevantes.

Entretanto, para que as métricas funcionem corretamente, é necessário que exista um conjunto de documentos classificados anteriormente, com o objetivo de identificar a quais consultas ou assuntos cada documento é relevante.

De acordo com Aggarwal e Zhai (2012), as principais métricas de avaliação de um SRI são: Precisão, Revocação, F-Score e Acurácia.

A precisão é uma métrica para analisar a quantidade de erros falsos positivos cometidos pelo classificador. Já a revocação mede a fração de exemplos verdadeiros positivos, ou seja, a fração de documentos classificados corretamente. A métrica F-score representa a média harmônica entre a precisão e a revocação. Por fim, a acurácia é a proporção de resultados classificados corretamente sobre o total de registros analisados pelo classificador (TAN; STEINBACH; KUMAR, 2005).

2.4.7 Técnicas de Análise de Coocorrência entre Termos em Documentos

De acordo com Sérgio, Silva e Gonçalves (2016), o termo correlação significa literalmente correlacionamento, sendo possível evidenciar o grau de relacionamento entre duas variáveis. O grau de correlação entre os termos contidos nos documentos textuais pode ser representado em cálculos oriundos da estatística. A finalidade do cálculo de correlação é a determinação da força do relacionamento entre dois elementos em análise (BARALIS et al., 2013). Entre os

modelos utilizados para determinar a correlação encontram-se a Frequência Conjunta, Média e Variância, Teste T (MANNING; MANNING; SCHÜTZE, 1999), o Chi-square (CHURCH; MERCER, 1993), o Phi-squared (CHURCH; GALE, 1991), a Informação Mútua (CHURCH; HANKS, 1990) e a Expected Mutual Information Measure (CHURCH; HANKS, 1990).

Neste estudo, a partir do corpus de treinamento oriundo do *dataset* anotado, de modo a obter o conjunto de termos representativos para cada aspecto, são aplicadas técnicas de análise de coocorrência entre termos nos documentos. De acordo com os estudos de Sápiras (2015), propõe-se a utilizar as técnicas Phi-squared (CHURCH; GALE, 1991) e Expected Mutual Information Measure - EMIM (CHURCH; HANKS, 1990). Estas técnicas são discutidas ao longo desta seção.

2.4.8 Expected Mutual Information Measure

A Expected Mutual Information Measure (EMIM), segundo Church e Hanks (1990), consiste em uma técnica para análise de coocorrência que compara a probabilidade de observar duas palavras, x e y , junto com a probabilidade de observá-las independentemente. Basicamente consiste em um modelo baseado em coocorrência entre termos, o qual emprega técnicas para realizar a agregação da informação.

EMIM é definida por meio da Equação 2.1, em que a é número de vezes que as palavras x e y co-ocorrem em um documento; b é o número de vezes em que x ocorre em um documento e y não ocorre; c é o número de vezes que y ocorre e x não ocorre em um documento e, d é o tamanho da coleção de documentos menos o número de documentos que não contenham x e/ou y , sendo que x representa uma palavra candidata e y representa uma palavra semente (anotadas por um especialista com base no conhecimento de domínio).

$$EMIM_{(x,y)} = \log_2 \frac{a(a + b + c + d)}{(a + b)(a + c)} \quad (2.1)$$

Caso o resultado de EMIM seja maior que zero, x é considerado termo representativo. Entretanto, de acordo com Sápiras (2015), experimentos preliminares demonstram resultados insatisfatórios utilizando apenas esta condição. Ainda de acordo com Sápiras (2015), buscando melhores resultados, pode-se mesclar o resultado de EMIM com o resultado da quantidade de vezes que x e y co-ocorrem em todos os documentos de determinado aspecto da base de

treinamento e através de experimentos, definir um número mínimo de coocorrências entre os dois termos para a seleção de x como termo representativo.

2.4.9 Phi-squared

A técnica phi-squared (ϕ^2) é uma medida estatística que favorece uma alta ocorrência de eventos (CHURCH; GALE, 1991). Phi-squared é definida por meio da Equação 2.2, cujas variáveis tem o mesmo significado das utilizadas na Equação 2.1.

$$\phi^2 = \frac{(ad - bc)^2}{(a + b)(a + c)(b + d)(c + d)}, \text{ onde: } 0 \leq \phi^2 \leq 1 \quad (2.2)$$

De maneira semelhante ao que pretende-se aplicar na abordagem EMIM, propõe-se a mesclar o resultado com Phi-Squared para análise dos experimentos, considerando a coocorrência entre os termos e os valores obtidos nas equações.

No próximo capítulo da presente dissertação, trabalhos relacionados são apresentados, com o intuito de ilustrar o estado da arte em estudos que utilizam mineração de dados no contexto de cidades inteligentes.

3 TRABALHOS RELACIONADOS

Este capítulo aborda a descrição da realização de estudos e desenvolvimento de trabalhos relacionados a pesquisas envolvendo mineração de dados e cidades inteligentes. Inicialmente é apresentada na seção 3.1 um breve resumo sobre o problema abordado por cada um dos trabalhos, procedimentos metodológicos e resultados obtidos. Em uma seção posterior 3.2, apresenta-se uma análise e discussão destes trabalhos quanto as limitações e contribuições.

3.1 ESTADO DA ARTE

3.1.1 Monitoramento de Mídias Sociais no Processo de Análise da Percepção dos Usuários Sobre a Segurança de Porto Alegre

Schaeffer (2016) propõe em seu trabalho, o estudo da percepção dos usuários sobre a segurança do município de Porto Alegre - RS, por meio de um estudo de caso no qual realiza a análise de dados extraídos do *Facebook* e *Twitter* entre 03 de março de 2016 a 30 de março de 2016. No caso do Facebook, foi considerada a página "Porto Alegre 24 Horas¹". Já no *Twitter*, por sua vez, foram considerados *tweets* (postagens) obtidos por combinações de palavras-chave associando o nome da cidade de Porto Alegre e a sigla POA com palavras relacionadas à segurança, insegurança, assalto, violência, entre outras. Além disso, considerou-se apenas publicações em português, uma vez que o *Twitter* permite que as buscas sejam filtradas pela linguagem declarada pelos seus usuários. As postagens foram coletadas pelo SentiMonitor², classificadas e categorizadas levando em consideração critérios como o conteúdo das postagens e o sentimento das mensagens. Para as publicações classificadas em Segurança, foi criado um conjunto de *tags*: assalto, protesto, violência, drogas, morte, furto-roubo, polícia-brigada, seg-geral. Para o estudo de caso extraiu-se 35.692 postagens. Destas, 18.980 foram classificadas em segurança, sendo que 12.360 são negativas, 2.820 neutras e 3.800 positivas.

¹ <https://www.facebook.com/portoalegre24horas>

² <https://www.sentimonitor.com>

3.1.2 Utilizando Redes Sociais para Detectar Crimes em Tempo Real

Neste trabalho, Figueredo e Sousa (2016) apresentam uma aplicação que utiliza uma rede social para suprir a necessidade de utilização de equipamentos de sensoriamento de elevado valor, com o objetivo de detectar crimes. Para este estudo, um sistema de aprendizagem de máquina *Naive Bayes* foi implementado para classificar mensagens oriundas do *Twitter*. Os dados foram captados a partir da API *Twitter4j*³, sendo que o classificador utilizou dados de ocorrências policiais para o treinamento. Além disso, na preparação dos dados, os *tweets* foram pré-processados a partir das técnicas de lematização e remoção de *stopwords*. Por fim, uma plataforma de processamento em tempo real foi implementada. Esta plataforma utiliza infraestrutura de processamento em tempo real como um componente central, implementada através do Apache Storm⁴. A estrutura desta plataforma pode ser visualizada na Figura 3.1.

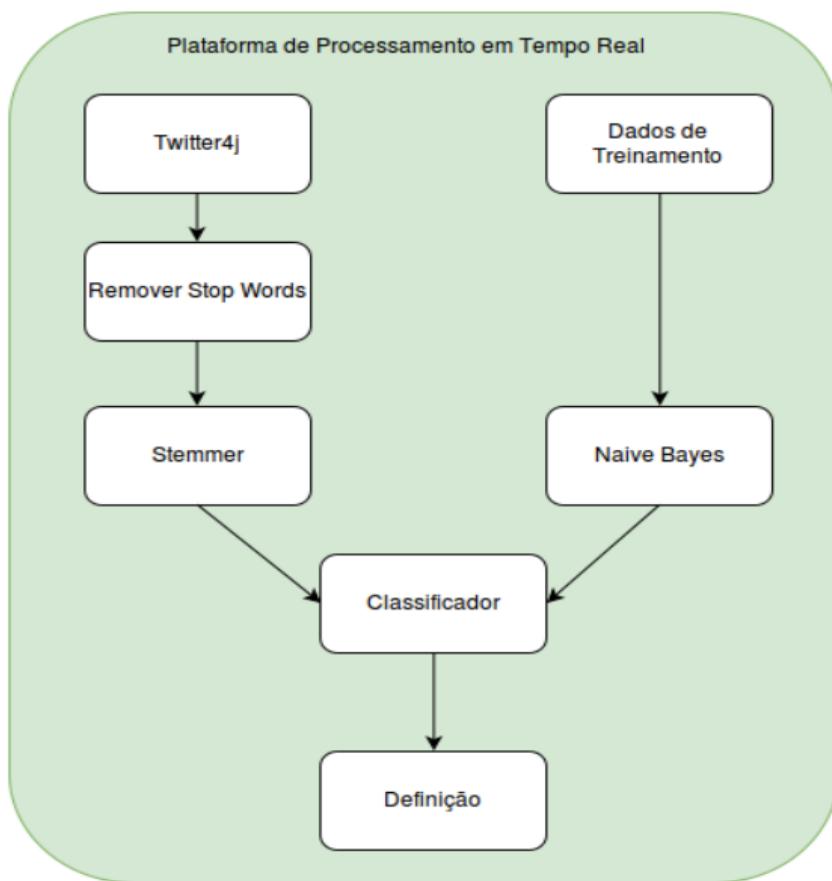


Figura 3.1: Plataforma para detecção de crimes em tempo real. Fonte: (FIGUEREDO; SOUSA, 2016).

Figueredo e Sousa (2016) destacam que esta plataforma de processamento em tempo

³ <http://twitter4j.org>

⁴ <http://storm.apache.org>

real absorveu uma boa quantidade de dados e foi capaz de processar todos, sem presença de erros. A ferramenta foi capaz de processar cerca de 2 tweets por segundo. Por outro lado, o classificador teve um bom desempenho com dados reais, porém houve um certa dificuldade quando tratou textos com muita ambiguidade ou teor cômico. Dentre os resultados desta pesquisa, o classificador *Naive Bayes*, considerando as técnicas de lematização e remoção de *stop words*, obteve 98% de acertos nas classificações.

3.1.3 Análise de Redes Sociais com Uso de Aprendizado de Máquina para Prever o Tráfego de Veículos em Zonas Urbanas

Nesta pesquisa, Dias (2017) propõe o estudo e desenvolvimento de técnicas de mineração de textos e aprendizado de máquina capazes de relacionar dados urbanos de fontes heterogêneas. O objetivo deste projeto é desenvolver e implementar um método em algoritmos de aprendizado de máquina capaz de identificar automaticamente se um *tweet* está informando sobre o trânsito e, caso o esteja, a que tipo de evento se refere. Quanto ao estudo de caso, em um período de duas semanas, na primeira coleta obteve-se um total de 36 GB de tweets, 690 GB em dados de fluxo e tráfego, 3.5GB para incidentes e 6.3 GB de dados meteorológicos. Já na segunda coleta, 3.4 GB de tweets, 235 GB de fluxo e tráfego e, 42 GB para incidentes e 7.1 GB de dados meteorológicos. Ambas as bases de dados foram filtradas, formatadas e logo após, unidas. No resultado desta fusão, foram empregadas técnicas de processamento de linguagem natural, com objetivo de preparar os dados para alimentar as rotinas de aprendizado de máquina implementadas. Ao final deste processo, obteve-se cerca de 300 Gigabytes de eventos de tráfego de veículos na cidade de São Paulo, e 10 Gigabytes de publicações do *Twitter* geolocalizadas na mesma cidade. O processo deste trabalho foi dividido em duas etapas: Coleta de dados e Análise dos Dados coletados. Na primeira etapa, foram coletados dados de trânsito, de meteorologia e de postagens do *Twitter*, utilizando as respectivas API's:

- *Traffic Flow API*⁵ e *Traffic Incidents API*⁶: Com a *Traffic Flow API* foram obtidos dados da fluência do trânsito no estado de São Paulo. Já a *Traffic Incidents API* foi utilizada para obter dados dos incidentes de trânsito no estado de São Paulo. Limite de 100.000 transações por mês.

⁵ https://developer.here.com/documentation/traffic/topics_v6.1/flow.html

⁶ <https://developer.tomtom.com/online-traffic/online-traffic-documentation/online-traffic-incidents>

- *Current Weather Data API*⁷: Foi utilizada para obtenção de dados meteorológicos (temperatura, pressão, velocidade do vento, entre outros) das cidades da região metropolitana da cidade de São Paulo. Limite de 60 requisições por minuto.
- *Streaming API*⁸: Disponibilizada pelo *Twitter* para aplicações que mineram grandes quantidades de dados continuamente. Foi utilizada para extrair os *tweets* publicados na região do estado de São Paulo. Limite não se aplica.

Inicialmente, foram coletados dados da região correspondente ao estado de São Paulo. Em seguida, dados que correspondem à região metropolitana da cidade de São Paulo. Estes dados extraídos a partir das API's foram pré-processados. Após o pré-processamento, durante a etapa de rotulação, o papel dos dados de trânsito foi servir de anotação para os *tweets*. Para isso foi realizada uma fusão dos *datasets* de *tweets* e incidentes de trânsito. Cada *tweet* foi relacionado com o incidente ao qual melhor corresponde, gerando assim um treinamento para o classificador. O período da primeira coleta foi de 28/04/2017 a 09/05/2017. Já o segundo período, de 07/06/2017 a 21/06/2017. Na implementação deste projeto, um script na linguagem de programação *Python* foi desenvolvido para a coleta de cada fonte de dados, todos construídos em cima de uma classe geral de mineração, também desenvolvida.

Na segunda etapa deste trabalho, no processamento de linguagem natural, foram aplicadas técnicas de *tokenização* nos dados coletados do *Twitter*, com o intuito capacitar o classificador a identificar se o texto analisado contém um (ou parte de um) endereço. Ademais, para identificar se está no começo, meio ou fim do texto. As páginas do *Twitter* utilizadas para extrair os dados foram: radiotransitofm, ZeroTransitoSP e TaxiAlphaville. Ao final da extração, obteve-se 1000 postagens de cada página, totalizando 3000 postagens.

Para a rotulação manual dos endereços no conjunto de dados foi utilizada uma expressão regular (DIAS, 2017). Após identificar ou não a presença de um endereço em cada um dos *tweets* coletados, a próxima etapa da análise consiste em aplicar um método de similaridade de palavras, a partir do conjunto de dados de incidentes de trânsito. Dessa forma, através de um método de similaridade entre o texto, endereço e descrição detalhada, é possível descrever se um *tweet* diz respeito a um incidente ou não. Para calcular a similaridade, foi usada a distância *Levenshtein*, também conhecida como distância de edição (LEVENSHTEIN, 1966). Já

⁷ <https://openweathermap.org/current>

⁸ <https://developer.twitter.com/en/docs/tweets/filter-realtime/overview.html>

a classificação dos incidentes relatados nos *tweets* foi realizada a partir de dois métodos, com o intuito de compará-los. A primeira técnica utiliza uma rede neural com uma camada de convolução, seguida de uma de *max-pooling* e uma totalmente conectada (*fully-connected*). Já a segunda técnica é uma Máquina de Vetor de Suporte (*SVM - Support Vector Machine*). Dentre os resultados, para visualizar a distribuição dos incidentes no mapa foi implementado um *script* que transforma uma tabela, na qual cada evento ocupa uma linha, para outra, em que as linhas representam a longitude e as colunas a latitude, e cada célula contém a soma de eventos que se encaixam naquela longitude e latitude. Diante desta nova tabela de frequências, utilizou-se a biblioteca *Seaborn*⁹ para transformá-la em um *heatmap*. Para melhor localização visual, foi feito um *heatmap* por cima da imagem do mapa usando a biblioteca *Folium*¹⁰. Para exemplo, a Figura 3.2 apresenta um *heatmap* dos incidentes que ocorreram às 07:43 do dia 12/06/2017, segunda-feira.

Com o intuito de melhorar a visualização no sentido de conseguir informar que tipo de evento aconteceu e onde, foi feito um outro mapa com a mesma biblioteca usando marcadores ao invés de um *heatmap*. A cor do marcador identifica a qual classe de incidentes ele pertence, sendo possível clicar no marcador para que uma caixa com mais informações seja apresentada. Este exemplo pode ser visualizado na Figura 3.3. Diante destas informações estruturadas, é possível realizar diversas análises e estudos, por períodos de intervalos determinados. Dias (2017) conclui que o objetivo desta pesquisa foi parcialmente cumprido, já que foi coletado um grande conjunto de dados de três fontes diferentes - *Twitter*, Trânsito e Meteorológico. Estes dados sofreram um intensivo tratamento e foram empregados no experimento de extração de endereços dos textos dos *tweets*, sendo implementadas algumas técnicas de aprendizado de máquina e processamento de linguagem natural. Além disso, foram produzidas visualizações que enriquecem a análise destes dados coletados.

⁹ <https://seaborn.pydata.org>

¹⁰ <https://python-visualization.github.io/folium/docs-master>

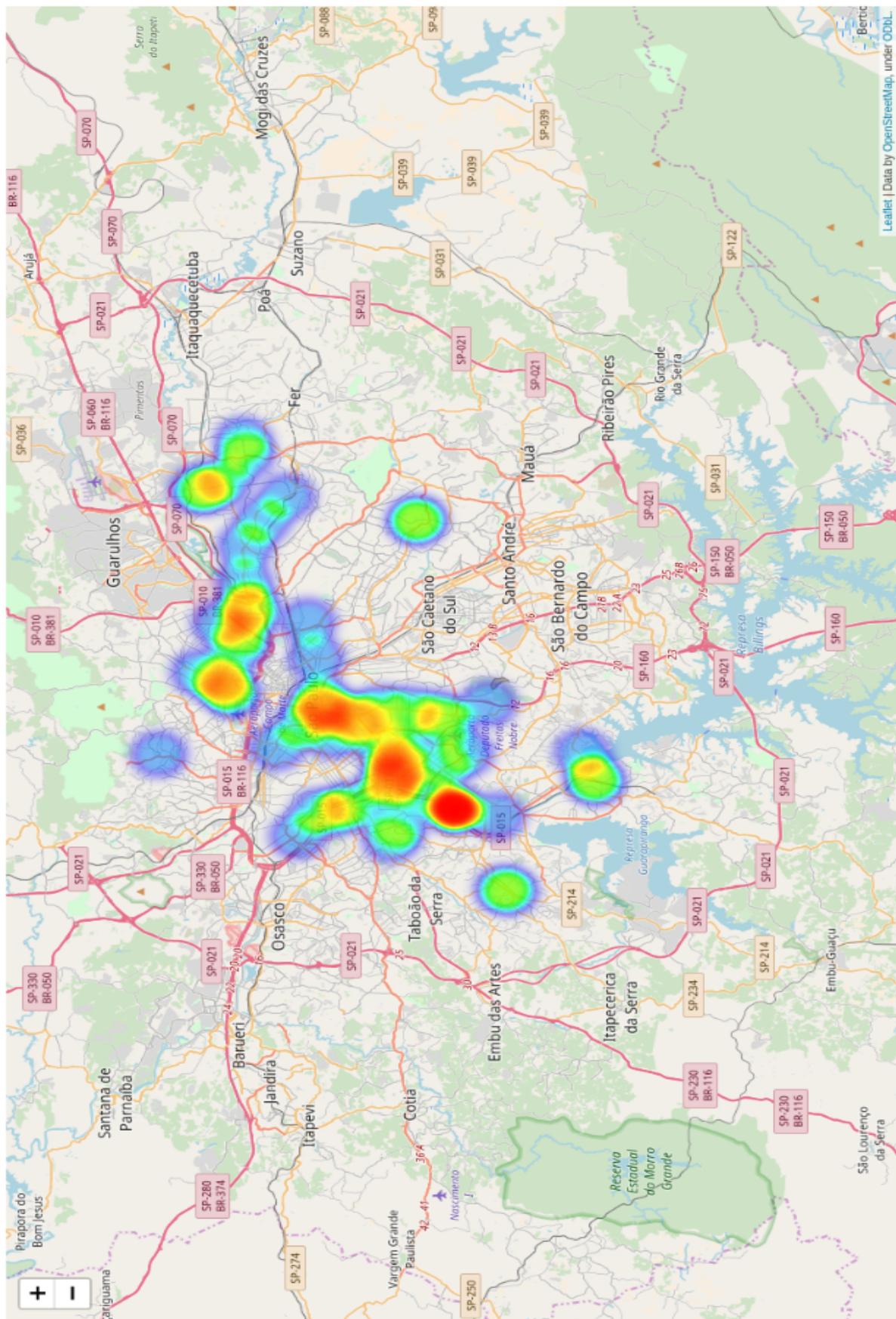


Figura 3.2: Heatmap dos incidentes que ocorreram às 07:43:00 no dia 12/06/2017 - Segunda-Feira. Fonte: (DIAS, 2017).

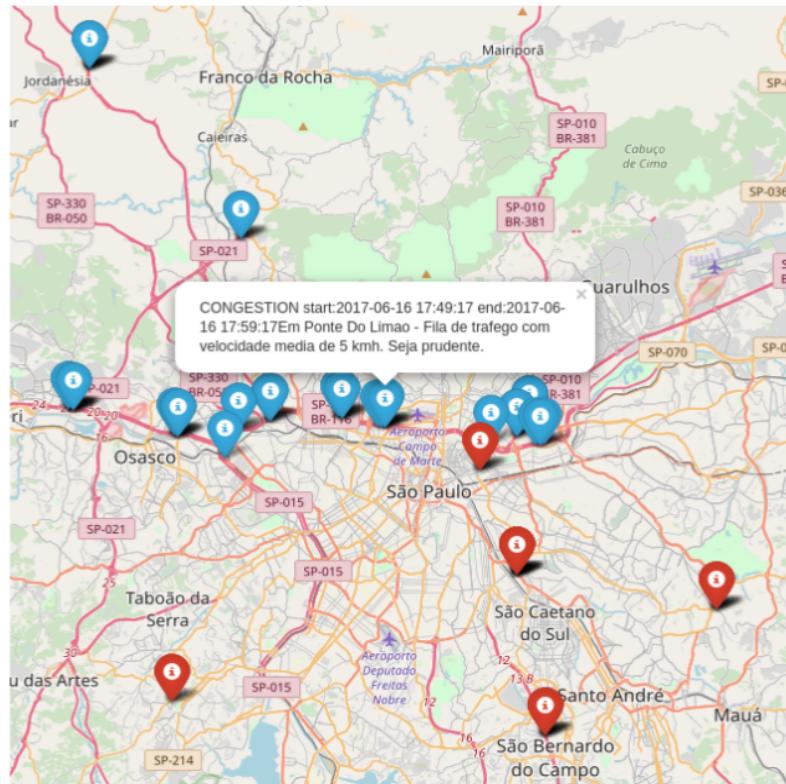


Figura 3.3: Mapa com marcadores dos Incidentes das classes Acidente e Congestionamento às 17:42:04 no dia 12/06/2017 - Segunda-Feira. Com um marcador de congestionamento selecionado. Fonte: (DIAS, 2017).

3.1.4 Uma Ferramenta de Rede Social Baseada em Localização Inspirada em Soluções de Cidades Inteligentes

Veloso (2012) sugere neste trabalho, uma aplicação de rede social baseada em localização, denominada *GeoPantIn*, implementada para o Pantanal no contexto de Cidades Inteligentes. Esta aplicação permite que cidadãos, órgãos governamentais e não governamentais, possam cadastrar denúncias sobre agressões a esta reserva da biosfera através de dados multimídia (fotos, vídeos, textos e áudio), bem como comentar *posts* já cadastrados, por meio de uma ferramenta de geoprocessamento acessível em navegadores Web de *desktops*, *smartphones* e *tablets*. O objetivo desta aplicação é criar um modelo de Pantanal Inteligente, a partir do desenvolvimento de um sistema georreferenciado, voltado ao cenário e às necessidades do Pantanal. Neste sistema, na tela principal, o mapa do Pantanal com as denúncias realizadas pelos usuários é exibido, possibilitando diversas opções de filtro por conteúdo (tipo de ocorrência, por período de tempo ou localização geográfica). Cada denúncia é apresentada no mapa na forma de marcador, contendo todos os dados informados pelo usuário no momento de seu cadastro, tais como

data do registro, localização e, se for o caso, algum arquivo multimídia anexado. Além disso, é possível realizar filtros para que sejam exibidos no mapa apenas os marcadores de interesse. Portanto, o sistema torna possível, por exemplo, visualizar as denúncias do tipo “Poluição” ou “Incêndio” que foram realizadas no ano de 2012 a uma distância máxima de 500m de um rio específico. Todas as informações do sistema são obtidas e apresentadas através de técnicas de recuperação de informações. O propósito deste trabalho consiste em auxiliar na manutenção do Pantanal e ajudar a contornar os problemas encontrados, buscando evitar incidentes mais graves, além de permitir que as pessoas participem mais ativamente na preservação de um dos mais importantes patrimônios ecológicos do nosso planeta. A partir dos dados cadastrados no sistema, os problemas ficam registrados e acessíveis em um meio público, podendo sofrer uma repercussão maior. Diante deste cenário, a sociedade pode realizar maior cobrança das autoridades. Em trabalhos futuros, pode ser realizada uma mineração de dados a partir dos dados cadastrados, de modo a encontrar padrões nos registros.

3.1.5 Priorização de Nós de uma Rede de Sensores sem Fio com Base na Análise de Dados de Redes Sociais

A conexão de sensores em forma de uma Rede de Sensores Sem Fio (RSSF) possibilita a criação de infraestruturas distribuídas para monitorar as cidades modernas. Nesse cenário, os sensores são posicionados criteriosamente para que cubram áreas bem definidas, como rodovias, estações de metrô e praças públicas. Neste contexto, a otimização na transmissão dos dados coletados possui um caráter extremamente importante no cenário das cidades inteligentes, sendo este um desafio ao grande tráfego de informações na rede, uma vez que centenas ou até mesmo milhares de sensores podem estar transmitindo informações simultaneamente (ANDRADE, 2017).

Neste trabalho Andrade (2017) propõe um sistema externo para RSSF a ser utilizado em cidades inteligentes. O objetivo principal deste sistema é otimizar uma RSSF para alta performance, a partir da detecção e classificação de eventos de interesse em mídias sociais para os nós desta rede, atribuindo níveis de relevância (na forma de índices de prioridade). O processo completo envolve 3 elementos distintos (Detector de Eventos, Classificador de Eventos e Protocolo de Definição de Prioridades), apresentados na Figura 3.4.

O primeiro módulo, Detector de Eventos, faz parte da Unidade de Cálculo de Rele-

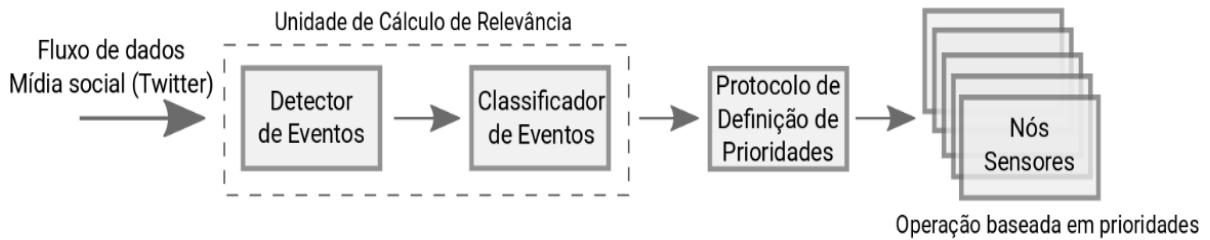


Figura 3.4: Esquema geral da abordagem proposta. Fonte: (ANDRADE, 2017).

vância, responsável por acessar o fluxo de dados de uma mídia social pré-definida (*Twitter*) em tempo real e procurar por eventos relevantes para a aplicação. Estes eventos, por sua vez, são submetidos ao Classificador de Eventos, que estabelece uma única prioridade numérica ao evento detectado. O último módulo deste sistema, consite no Protocolo de Definição de Prioridades, que conta com a Unidade de Cálculo de Relevância, a qual está completamente operacional. Entretanto os algoritmos de configuração dos sensores para a definição das prioridades ainda estão sob desenvolvimento e serão abordadas em trabalhos futuros. Quanto ao estudo de caso e resultados, a proposta foi aplicada utilizando um *dataset* real composto por 1.4 milhões de *tweets* obtido através da API pública de *streaming* do *Twitter*. Esses dados representam uma fração de todos os *tweets* geolocalizados publicados entre 07/02/2017 e 30/04/2017 na cidade de Nova York - EUA. O *dataset* foi processado para verificar os procedimentos de assimilação da prioridade contextual. O algoritmo foi executado em janelas de tempo de uma hora e em sua configuração padrão (ZHANG et al., 2016). A Figura 3.5 apresenta a quantidade média de cada tipo de evento por dia da semana.

Durante os testes mencionados, a Unidade de Cálculo de Relevância classificou corretamente mais de 70% dos eventos e, utilizando os resultados obtidos, é possível identificar alguns padrões, como por exemplo, que a quantidade de eventos relativos à mobilidade dobra durante os sábados e domingos, indicando que a população enfrenta problemas notáveis relacionados ao tráfego ao decorrer desses dias. Utilizando esses *insights*, os sistemas da cidade podem ser otimizados para dar suporte a tais eventos, diminuindo o impacto de eventos correntes e de eventos futuros. Os resultados obtidos mostram que a abordagem proposta pode trazer resultados bastante significativos para o monitoramento de cidades inteligentes utilizando RSSF, considerando que informações extraídas da Internet possam ser inseridas permitindo diversas estratégias de otimização na rede.

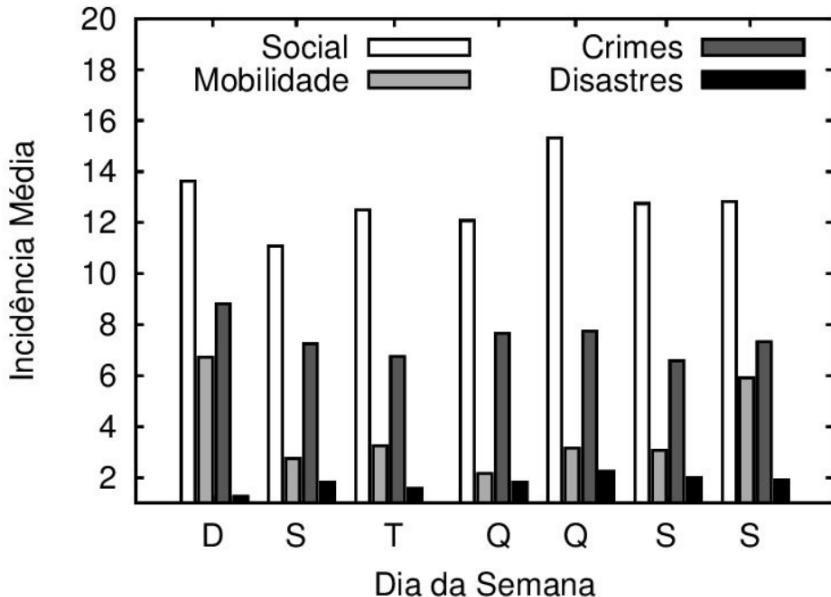


Figura 3.5: Eventos detectados no *dataset* de Nova York. Fonte: (ANDRADE, 2017).

3.2 ANÁLISE COMPARATIVA

A Tabela 3.1 mostra a relação entre os estudos que abordam mineração de dados e cidades inteligente citados anteriormente e a presente pesquisa.

A coluna **Base de dados**, informa a fonte em que foram coletados os dados para análise. A coluna **Período de coleta dos dados** diz respeito ao intervalo de período inicial e final referente à coleta dos dados. Na coluna **Amostra de dados**, são apresentadas as quantidades de dados coletados para os estudos, sendo que alguns trabalhos mencionam dados números relacionados a *posts* do *Facebook* ou *tweets* do *Twitter*, enquanto que outros informam unidade de medida de informação em bytes. A coluna **Aspecto(s) analisado(s)** está relacionada aos assuntos ou temas abordados em cada pesquisa. Já a coluna **Métodos computacionais aplicados** refere-se aos algoritmos ou técnicas de mineração aplicadas em cada pesquisa. Por fim, a coluna **Cidade** informa o município alvo de cada trabalho.

Tabela 3.1: Tabela Comparativa dos Trabalhos Relacionados.

Trabalho	Base de dados	Período de coleta dos dados	Amostra de dados	Aspecto(s) analisado(s)	Métodos computacionais aplicados	Cidade
3.1.1	<i>Facebook e Twitter</i>	03/03/2016 a 30/03/2016	35692	Segurança	<i>Sentimonitor</i> (plataforma privada)	Porto Alegre - RS, Brasil
3.1.2	<i>Twitter</i>	Não informado	Não informado	Segurança	<i>Naive Bayes</i>	Não informado
3.1.3	- <i>Twitter</i> - <i>Traffic Flow</i> - <i>Traffic Incidents</i> - <i>Current Weather Data</i>	- 08/04/2017 a 09/05/2017 e - 07/06/2017 a 21/06/2017	- 300 GB em eventos de tráfego e meteorologia - 10 GB de tweets	Mobilidade Urbana	Redes neurais e Máquina de Vetor de Suporte (<i>SVM - Support Vector Machine</i>)	São Paulo - SP e região metropolitana, Brasil
3.1.4	Própria	Não informado	No informado	Meio Ambiente	Recuperação de Informações	Pantanal - Amazonia, Brasil
3.1.5	<i>Twitter</i>	07/02/2017 a 30/04/2017	1.4 milhões de tweets	Sociais, Mobilidade, Segurança, Meio Ambiente	Processamento de linguagem natural e agrupamento de dados através da ferramenta <i>GeoBurst</i> e princípio da máxima a posteriori	Nova York - Nova York, EUA
Presente Dissertação	<i>Facebook</i>	18/06/2011 a 10/07/2017	53579	Economia, Educação, Empreendedorismo, Energia, Governação, Meio Ambiente, Mobilidade, Segurança, Saúde, Tecnologia & Inovação e Urbanismo	<i>Expected Mutual Information Measure e Phi-Squared</i>	Rio Grande - RS, Brasil

Diante dos trabalhos apresentados na Tabela 3.1, observa-se que as pesquisas foram aplicadas para um curto intervalo de tempo. Em relação aos aspecto(s) analisado(s), percebe-se que maioria dos trabalhos aplicam a metodologia considerando somente um assunto ou tema. Quanto as bases de dados mencionadas em cada estudo, o *Twitter* foi a rede mais utilizada, porém, mesmo sendo uma potente fonte para obtenção de grandes volumes de dados, está limitada para apenas 280 caracteres em cada publicação. Além disso, no Brasil o *Twitter* é menos utilizado do que o *Facebook* (STATISTA, 2017b). Por fim, em relação aos métodos computacionais aplicados, todos trabalhos apresentam bons resultados. Entretanto, somente a pesquisa que envolve a cidade de Nova York aplicou as técnicas para mais de uma classe (assunto ou tema).

O capítulo a seguir da presente dissertação, apresenta a metodologia aplicada neste estudo e discute os detalhes de implementação, tais como as etapas de pré-processamento, ferramentas computacionais utilizadas e técnicas estatísticas aplicadas para o processo de mineração de textos.

4 UM MODELO PARA MINERAÇÃO DE DADOS DE POSTAGENS DO FACEBOOK PARA INFERÊNCIA DA INTELIGÊNCIA DO TERRITÓRIO

Neste trabalho busca-se estudar técnicas e ferramentas de mineração de dados, capazes de permitir inferir o grau de inteligência de um território, a partir de postagens do Facebook. De forma mais precisa, apresenta-se um modelo e suas ferramentas capazes de pré-processar estas postagens, bem como fornecer termos representativos associados aos diferentes aspectos identificados na literatura como importantes na determinação do nível de inteligência de uma cidade. A partir dos termos representativos, propõe-se também prover um método capaz de classificar uma postagem quanto a um aspecto. Os termos representativos são obtidos com base no fluxograma apresentado na Figura 4.1, através das técnicas EMIM e Phi-Squared, considerando como fontes de entrada o corpus anotado de cada aspecto, além das palavras sementes e candidatas. Já o fluxograma ilustrado na Figura 4.2, apresenta a sequência do conjunto de ferramentas utilizadas neste modelo. As etapas desse processo são descritas ao longo desta seção.

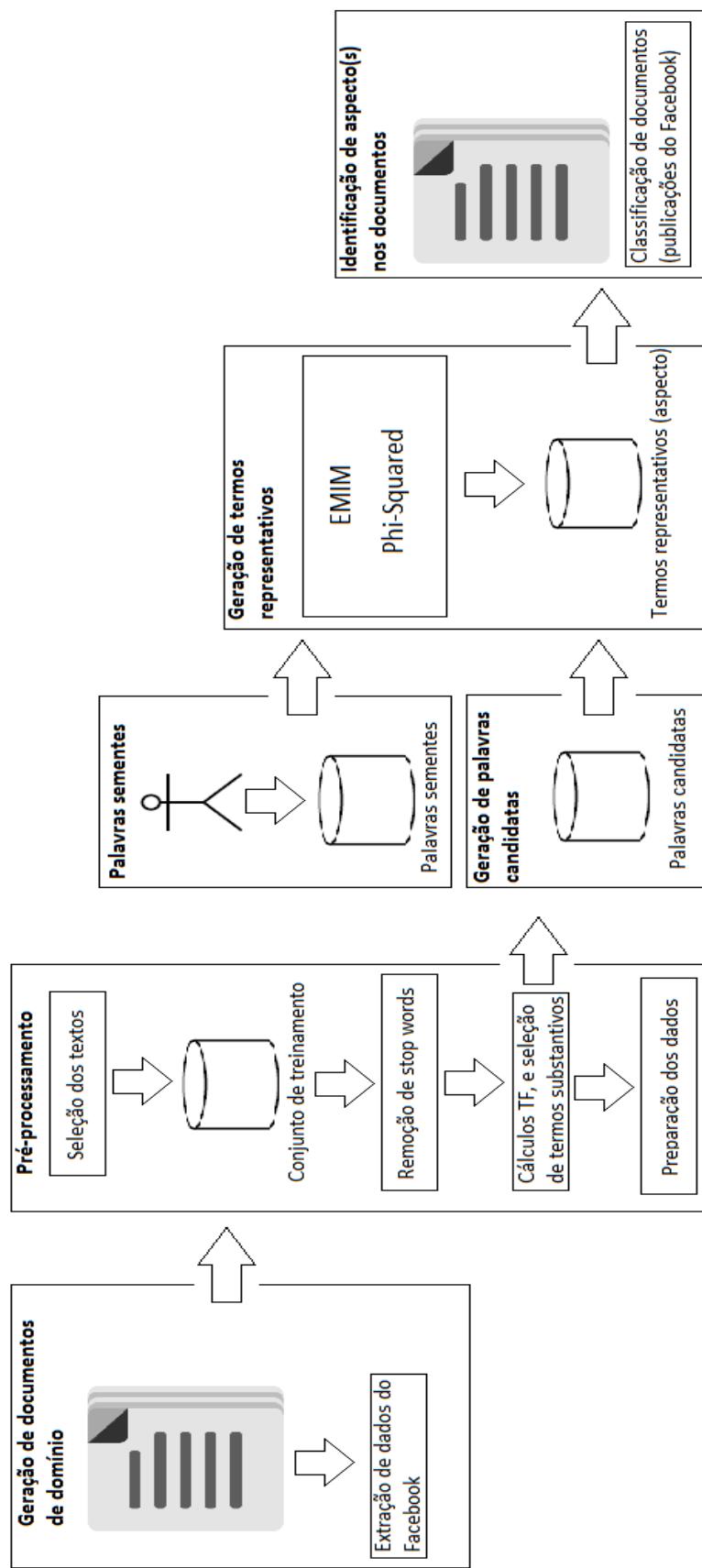


Figura 4.1: Fluxograma da abordagem proposta.

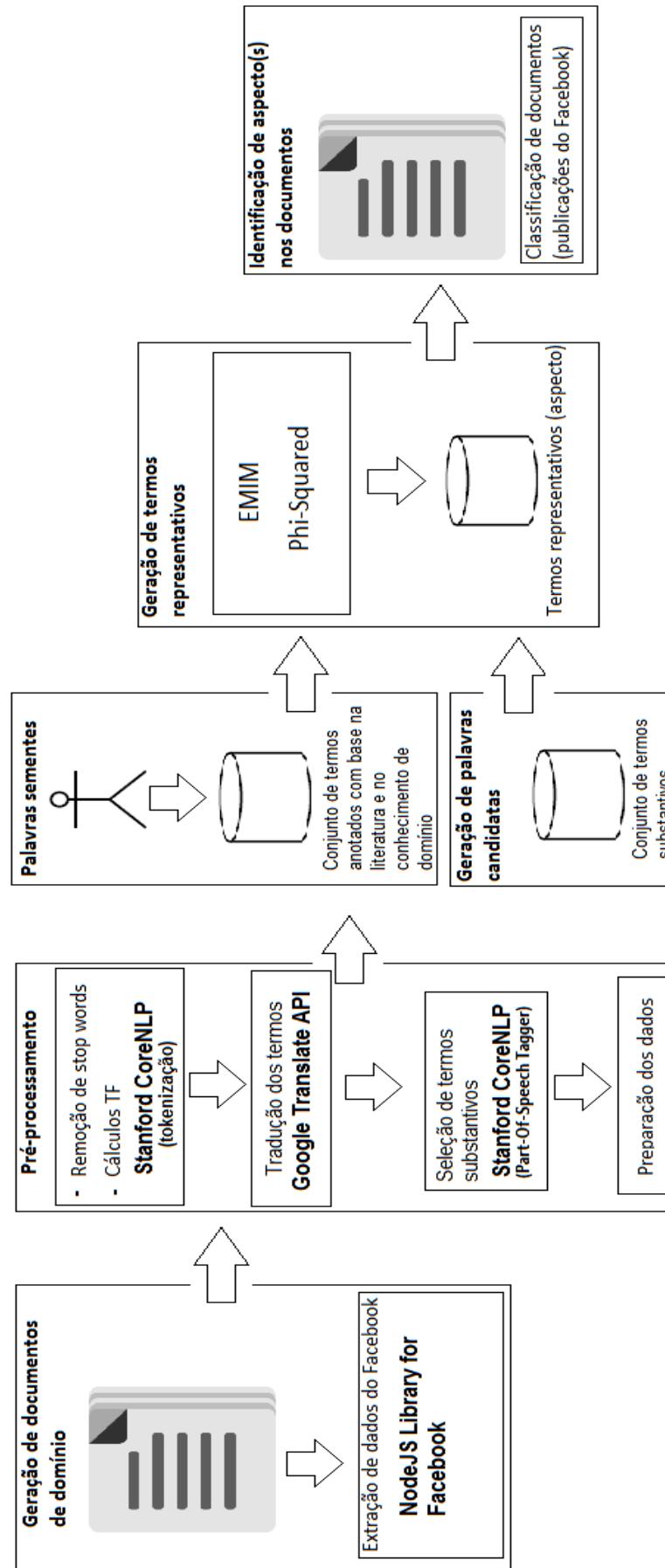


Figura 4.2: Fluxograma da abordagem proposta.

4.1 GERAÇÃO DE DOCUMENTOS DE DOMÍNIO

De acordo com os objetivos apresentados nesta dissertação, com o intuito de obter o conjunto de dados de domínio, propõe-se selecionar páginas e grupos temáticos da rede social Facebook para a extração dos dados, sendo estas, públicas e privadas.

Para a extração dos dados, sugere-se a ferramenta NodeJS Library for Facebook¹, a qual trata-se de uma implementação cliente em NodeJS para o serviço Web da Graph API Explorer, disponibilizado pelo Facebook para extração de dados da rede social. Vale destacar que faz-se necessário criar uma conta na página de desenvolvedores do Facebook, de modo a gerar um *token* de acesso para ser utilizado na aplicação cliente.

Neste contexto, extrai-se os dados em arquivos no formato JSON, que são migrados para a base de dados objeto-relacional MySQL². Para o processo de migração dos dados no formato JSON ao MySQL, propõe-se a utilização da linguagem de programação Java, na plataforma de desenvolvimento JavaSE, junto ao JPA Hibernate para persistência na base de dados, bem como a utilização de bibliotecas para manipulação dos arquivos no formato JSON. O resultado deste processo é a criação e população das tabelas da base de dados, ilustrado na Figura 4.3 através de um diagrama do modelo objeto-relacional.

¹ <https://www.npmjs.com/package/fb>

² <https://www.mysql.com>

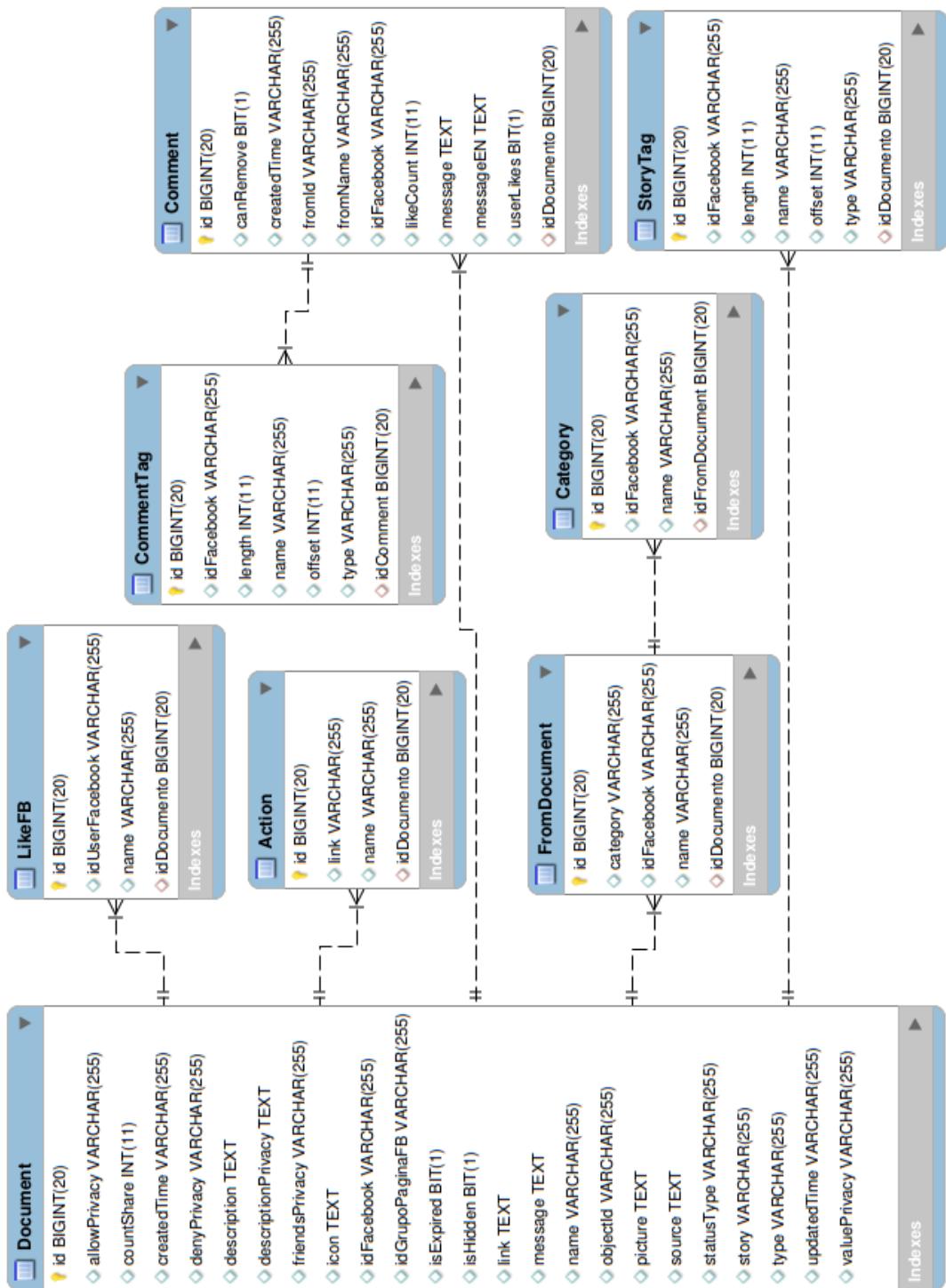


Figura 4.3: Diagrama do modelo objeto-relacional da base de dados.

4.2 PRÉ-PROCESSAMENTO

Em um processo de descoberta de conhecimento – KDD, a etapa de pré-processamento compreende a aplicação de várias técnicas para captação, organização, tratamento e preparação dos dados. Possui fundamental relevância no processo de KDD, já que contempla desde a correção de dados incorretos até o ajuste da formatação dos dados para os algoritmos de mineração de dados que serão aplicados.

Nesta proposta, o pré-processamento dividide-se em seis (6) etapas: seleção dos textos, geração do conjunto de treinamento, remoção de *stop words* e termos simples, cálculos de frequência dos termos, anotação de termos substantivos e preparação dos dados, as quais são apresentadas ao longo desta seção para a geração de palavras candidatas e termos representativos.

4.2.1 Seleção dos textos

Considerando que o pré-processamento dos dados abrange qualquer tratamento realizado sobre os dados selecionados de forma a assegurar a qualidade (completude, veracidade e integridade) dos fatos por eles representados, informações ausentes, incorretas ou inconsistentes, devem ser corrigidas de forma a não comprometer a qualidade dos modelos de conhecimento a serem extraídos ao final do processo de KDD. Nesta proposta, o pré-processamento dos dados consiste primeiramente, na seleção dos textos, considerando o tratamento e preparação dos dados para a geração do conjunto de treinamento (anotação manual dos documentos em aspecto(s)).

4.2.2 Geração do conjunto de treinamento

Para esta etapa propõe-se o desenvolvimento de um sistema Web, utilizando a linguagem de programação PHP³, de modo a permitir que diferentes usuários classifiquem ou anotem os documentos em aspecto(s), gerando assim um *dataset* anotado. Neste sistema, cada publicação apresentada pode ou não ser anotada em um ou n aspectos. As publicações que não são anotadas ou classificadas em n aspectos, são desconsideradas para a geração do *dataset* anotado, pois referem-se a mais de um dos aspectos ou a nenhum.

³ <http://php.net>

As telas do sistema web proposto são ilustradas através das Figuras A.1, A.2, A.3, A.4 e A.5, localizadas no apêndice deste documento.

É importante destacar, que este sistema web apresenta de maneira aleatória, uma publicação a cada iteração. Entretanto, contempla lógica na consulta a base de dados para priorizar a apresentação e anotação de cada documento por 5 usuários diferentes do sistema, com o intuito de obter um senso comum. A Figura A.6, localizada no apêndice deste trabalho, ilustra a consulta (*query*) implementada para apresentação das publicações.

A Figura 4.4 ilustra a tabela *DocumentClassification* da base de dados MySQL onde são armazenadas as anotações dos documentos, bem como as tabelas relacionadas *Document* e *UserClassificationWebSystem*, sendo esta última referente aos usuários do sistema.

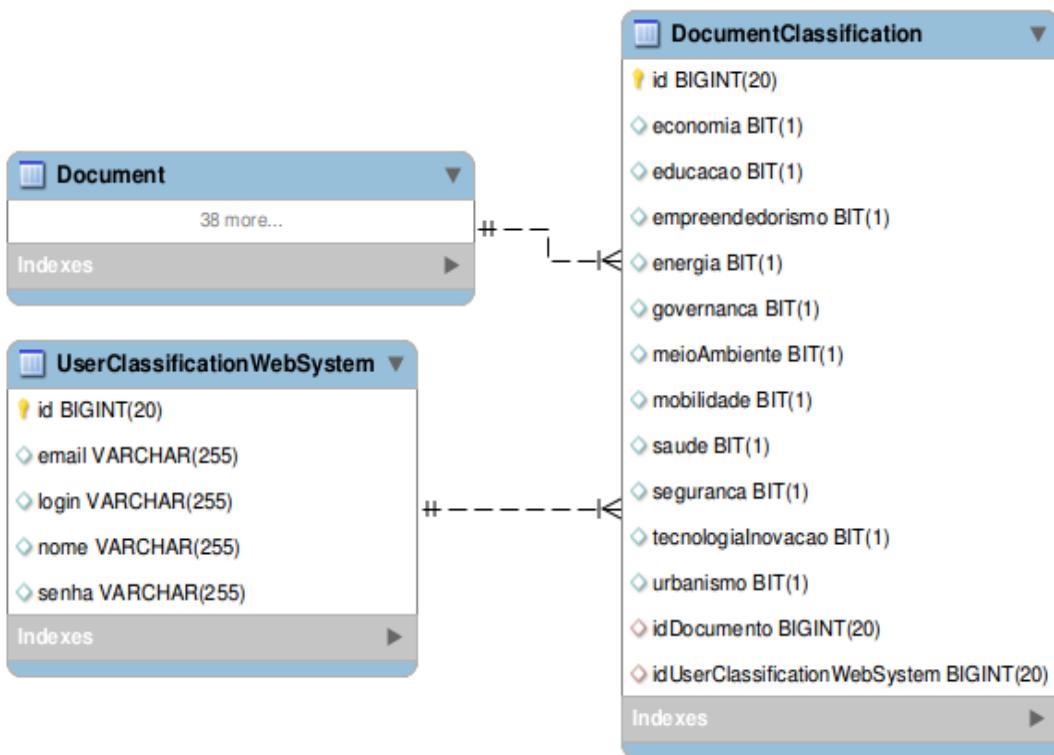


Figura 4.4: Tabela *DocumentClassification* para armazenamento das anotações dos documentos e tabelas relacionadas *Document* e *UserClassificationWebSystem*.

Diante das publicações anotadas a partir do sistema web, cada documento foi julgado em determinado aspecto quando teve 5, 4 ou 3 anotações, por usuários diferentes do sistema, observando a unanimidade ou a maioria absoluta dos votos em apenas um dos aspectos.

A Figura 4.5 apresenta as tabelas criadas e populadas para armazenar os documentos anotados nos respectivos aspectos, após julgados. É importante destacar que todas tabelas

com sufixo "Matriz" contêm os mesmos atributos. Estes atributos podem ser visualizados na tabela MatrizEconomia. O campo "treinamento" é criado para denominar se, a partir do *dataset* anotado, o documento é de validação ou treinamento, para os diferentes experimentos que são realizados nesta pesquisa. Os documentos de treinamento são utilizados para aplicar as técnicas de mineração deste trabalho, com o intuito de obter os termos representativos para cada aspecto. Já os documentos de validação são utilizados para classificação e avaliação da qualidade do classificador através das métricas de precisão, revocação, f-score e acurácia.

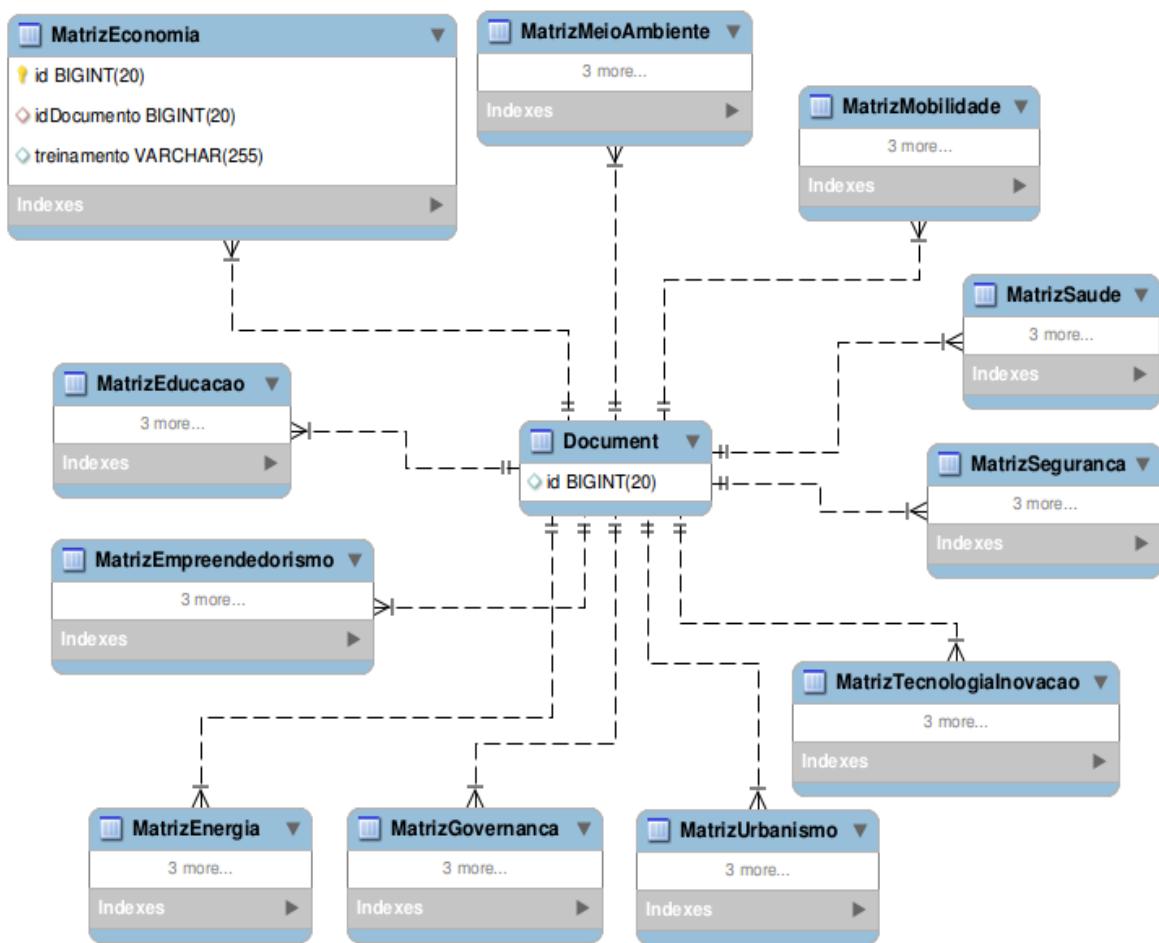


Figura 4.5: Tabelas criadas e populadas para armazenar os documentos anotados nos respectivos aspectos, após julgados por unânimidade ou maioria absoluta de votos no aspecto.

4.2.3 Remoção de *stop words* e termos simples

De posse do conjunto de treinamento para cada aspecto, apresentado através das tabelas ilustradas na Figura 4.5, propõe-se a identificação e remoção de termos simples presentes nos textos, tais como símbolos e caracteres especiais, bem como datas e números, além de *stop*

words, que são palavras de parada consideradas irrelevantes para o conjunto de treinamento, pois não carregam significado em seu contexto, tais como preposições e artigos de ligação. Vale observar que as *stop words* e os termos simples são removidos para realizar os cálculos de frequência dos termos, ou seja, elas não pertencem nos cálculos.

4.2.4 Cálculos de frequência dos termos

Após a identificação e remoção das *stop words* e termos simples, propõe-se a realização dos cálculos de frequência dos termos nos documentos, os quais são utilizados nas técnicas deste trabalho para o processo de mineração de textos. Para estes cálculos de frequência absoluta dos termos nos documentos, criou-se duas tabelas, *WordCloud* e *WordCloudTF*, conforme estão apresentadas através da Figura 4.6.

Primeiramente construiu-se um modelo booleano, através da criação e população da tabela *WordCloud*. Nesta etapa, propõe-se um processo de *tokenização* nos textos, que divide os mesmos em sentenças e termos. Para este processo recomenda-se a utilização da ferramenta Stanford CoreNLP⁴, que trata-se de um projeto desktop, desenvolvido pela Universidade de Stanford⁵, para o processamento de linguagem natural, composto por um conjunto de classes e bibliotecas disponíveis na linguagem de programação Java. É importante destacar que durante o processo de *tokenização*, para cada palavra encontrada nos textos do conjunto de treinamento, armazena-se a frequência da mesma no documento. Vale informar que os dados cadastrados na tabela *WordCloud* são oriundos dos documentos presentes nas tabelas apresentadas na Figura 4.5, onde o campo "treinamento" é igual a 1. Além disso, também é importante destacar que nesta tabela, *WordCloud*, se um termo ocorre duas vezes em determinado documento, dois registros são armazenados para esta informação, sendo estes idênticos. Uma vez que a tabela *WordCloud* contém todos os devidos registros, cria-se e popula-se a tabela *WordCloudTF*, a qual serve de base para o armazenamento das ocorrências dos termos nos documentos. Ademais, ela também armazena a classe gramatical das palavras e os termos traduzidos para o idioma inglês, os quais são utilizados na próxima etapa do processo, que consiste na anotação de todos os substantivos identificados no corpus de treinamento.

A criação e população da tabela *WordCloudTF* são ilustradas a partir das Figuras A.7 e A.8, localizadas no apêndice deste documento. Popula-se a tabela *WordCloudTF* através do

⁴ <https://stanfordnlp.github.io/CoreNLP>

⁵ <https://www.stanford.edu>

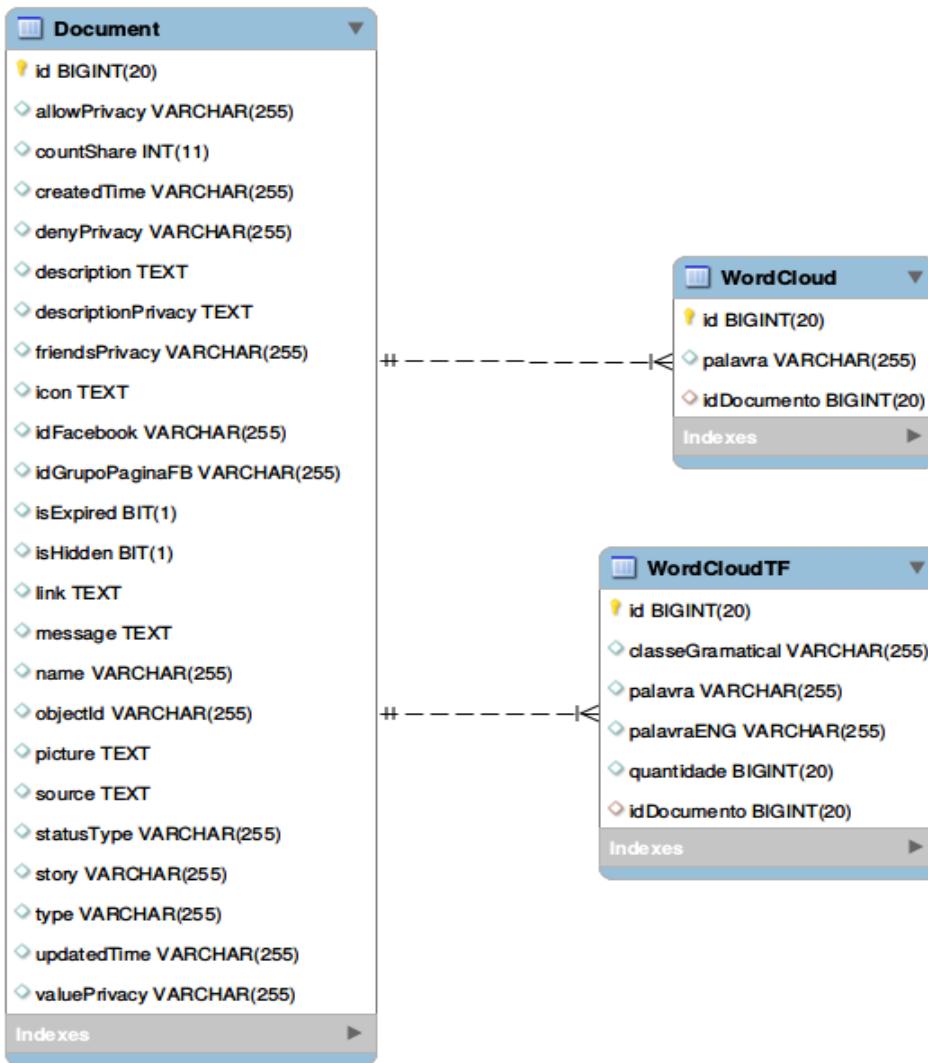


Figura 4.6: Tabelas *WordCloud* e *WordCloudTF* da base de dados MySQL, relacionadas à tabela *Document*, respectivamente para geração de um modelo booleano entre termos e documentos e armazenamento das frequências dos termos nos documentos.

agrupamento dos registros da tabela *WordCloud*, considerando a quantidade de ocorrências de cada palavra em determinado documento.

Após criar e popular a tabela *WordCloudTF*, os termos são traduzidos para o idioma inglês e armazenados no campo “palavraENG” desta tabela, através das API’s Google Translate e JPA Hibernate, respectivamente.

4.2.5 Anotação de termos substantivos

Com o intuito de gerar e obter um conjunto de palavras candidatas para cada aspecto, a partir da identificação, anotação e seleção de todos substantivos encontrados nos respecti-

vos corpus de treinamento, nesta etapa final de preparação dos dados, propõe-se um processo de *Part-Of-Speech Tagger*, utilizando a ferramenta Stanford CoreNLP. Este processo realiza a identificação de categorias das palavras ou propriedades gramaticais. Entretanto, os termos analisados devem ser traduzidos para o idioma inglês, devido a ausência de suporte desta ferramenta para o idioma português. Para a tradução dos termos, propõe-se a utilização da API (Application Programming Interface) Google Translator⁶.

4.2.6 Preparação dos dados

Após a seleção dos textos, geração do conjunto de treinamento, remoção de *stop words* e termos simples, cálculos de frequência dos termos e anotação dos substantivos nos corpus de treinamento de cada aspecto, a preparação dos dados está concluída para as próximas etapas, que consiste na geração de palavras candidatas, a partir dos substantivos identificados, além das anotações de palavras sementes, as quais ambas são fontes de entrada para a geração dos termos representativos, a partir das técnicas de mineração aplicadas neste trabalho.

4.3 PALAVRAS SEMENTES

As palavras sementes neste projeto são anotadas para cada aspecto com base na literatura e a partir do conhecimento de domínio. No próximo capítulo elas são apresentadas para os diferentes experimentos realizados. É importante observar que o conjunto de palavras semente é independente do conjunto de palavras candidatas.

4.4 GERAÇÃO DE PALAVRAS CANDIDATAS

Para a geração das palavras candidatas, a partir de um processo de *Part-Of-Speech Tagger* utilizando a ferramenta Stanford CoreNLP, analisou-se todos os termos traduzidos para o idioma inglês armazenados na tabela *WordCloudTF*. Quando identificou-se que o termo analisado trata-se de um substantivo, o mesmo foi anotado como palavra candidata, considerando o aspecto do documento de treinamento.

⁶ <https://translate.google.com>

4.5 GERAÇÃO DOS TERMOS REPRESENTATIVOS

Para este estudo, nesta etapa propõe-se a utilização de duas técnicas estatísticas de análise de coocorrência entre termos nos documentos: Phi-squared e Expected Mutual Information Measure. Ambas as técnicas, utilizam como entrada o conjunto de palavras sementes, além do conjunto de palavras candidatas e os documentos da base de treinamento.

Nesta pesquisa foram realizados dez experimentos diferentes. Para cada experimento utilizou-se uma base de dados individual, sendo que a partir do conjunto de dados de treinamento, considerou-se de maneira aleatória diferentes documentos para validação. Para os aspectos Empreendedorismo e Tecnologia & Inovação, foram separados 30% dos documentos anotados para validação, devido ao fato destes contemplarem poucas anotações. Já nos demais, foram considerados para validação 20% do *dataset* anotado. Na implementação das técnicas de análise de co-ocorencia entre termos nos documentos, EMIM e Phi-Squared, que são propostas neste trabalho, realizou-se a combinação entre cada documento de treinamento que não foi anotado para validação, com cada palavra semente e candidata.

Para cada iteração do laço de repetição entre documentos de treinamento, palavras sementes e palavras candidatas, realizou-se os cálculos das equações referentes às técnicas EMIM e Phi-Squared, de acordo com a literatura. Quando uma palavra candidata tornou-se termo representativo, ela foi armazenada em tabela da base de dados. Com o intuito de armazenar estes termos, criou-se uma tabela “EMIMPhiSquared” para cada aspecto, conforme ilustrado na Figura 4.7. Vale salientar que todas as tabelas contém os mesmos atributos, ilustrados na tabela “EMIMPhiSquaredEconomiaTermoDoc”.

4.6 IDENTIFICAÇÃO DE ASPECTOS NOS DOCUMENTOS

Após obter um conjunto de termos representativos para cada aspecto, a próxima etapa deste trabalho consiste em analisar os documentos de validação. Estes documentos variam entre um percentual de 20% a 30% dos documentos anotados, conforme descritos na seção anterior deste capítulo. Por fim, propõe-se a classificá-los de acordo com a identificação de termos representativos. Para a classificação de um documento em um aspecto, considera-se a maior quantidade encontrada de termos representativos de determinado aspecto, classificando o documento nesta classe. Em caso de empate, o documento não é classificado.

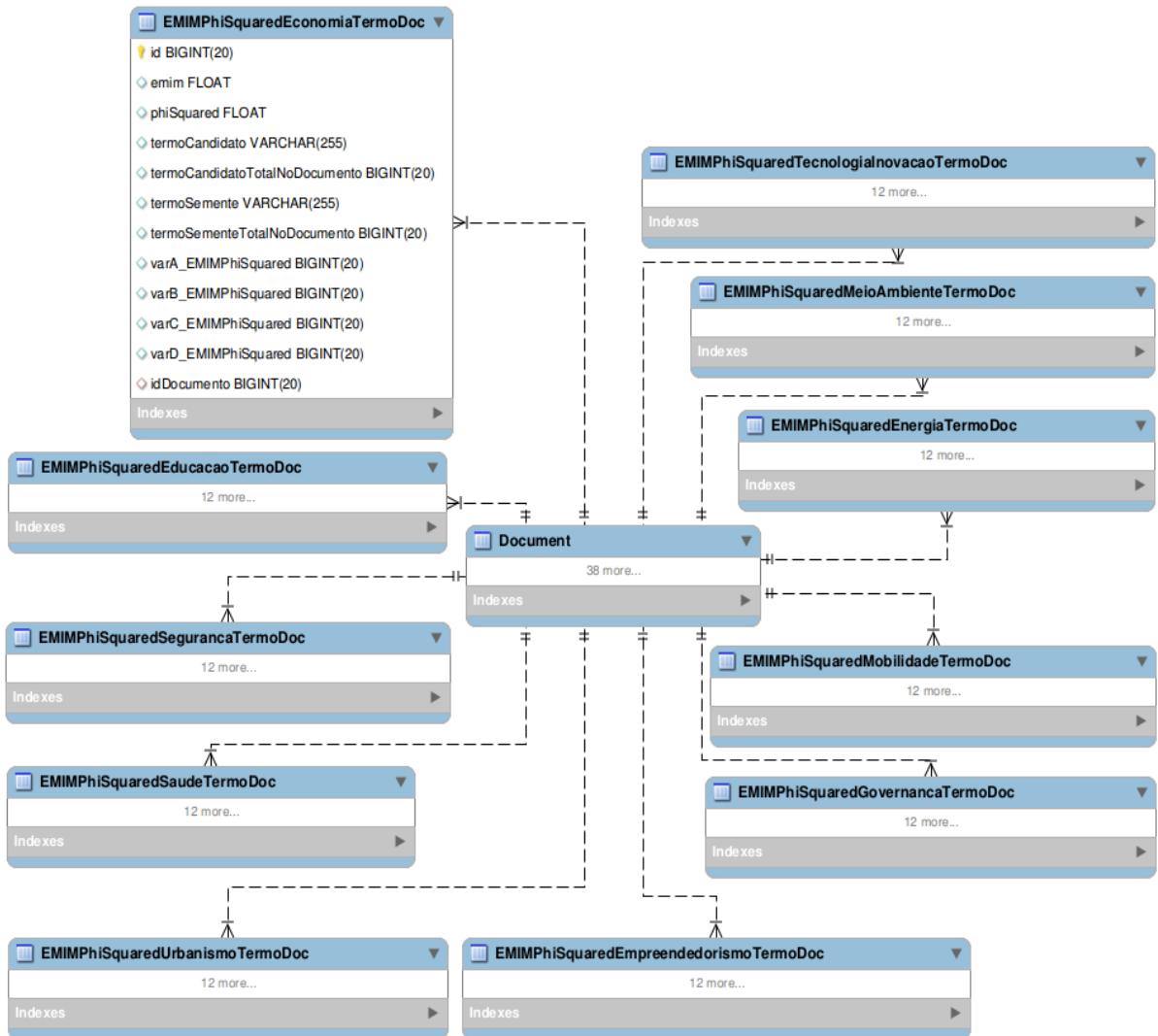


Figura 4.7: Tabelas EMIMPhiSquared para armazenamento de termos representativos.

Para realizar a avaliação dos resultados, referentes às técnicas de mineração de textos, de modo a obter um índice de qualidade sobre o classificador de documentos, neste trabalho são utilizadas as métricas de precisão, revocação, f-Score e acurácia. Após esta análise, propõe-se classificar todos os documentos da base de dados que não foram anotados durante o treinamento.

No próximo capítulo, o estudo de caso desta pesquisa é apresentado, o qual descreve os experimentos realizados, bem como os dados utilizados para a etapa de mineração e resultados obtidos.

5 TESTES E VALIDAÇÃO

5.1 DELIMITAÇÃO DO ESTUDO DE CASO

Esta seção apresenta, através das Tabelas 5.1 e 5.2, as páginas e grupos temáticos da rede social Facebook os quais foram extraídos dados textuais de publicações para análise, sendo estes relacionados às informações da cidade do Rio Grande/RS.

O período considerado para coleta de dados foi desde a fundação de cada grupo ou página, até 10 de julho de 2017, data em que expirou a versão de serviço 2.3 da API Graph Explorer. A versão 2.3 da Graph API foi considerada para estudo de caso devido ao fato de conter maior quantidade de informações nas publicações, comparada às versões superiores. Nas versões superiores, são disponibilizados apenas dados relacionados às publicações, contendo menos atributos. Além disso, na versão 2.3 também são acessíveis e retornados dados de comentários, curtidas (*likes*), ações permitidas sobre as publicações, tais como: compartilhar (*share*), curtir (*like*), comentar (*comment*) e denunciar postagem ao administrador (*report post to admin*); entre outros objetos com informações relevantes, também relacionados às publicações, os quais são ilustrados na Figura 4.3 e que podem ser utilizados em trabalhos futuros. No entanto, vale observar que neste trabalho são analisadas apenas as publicações.

Para a seleção dos textos, a partir dos documentos de domínio descritos no capítulo 3, considerou-se apenas documentos distintos que obrigatoriamente contêm textos nas publicações, com no mínimo 50 caracteres, desconsiderando publicações que contêm apenas fotos, vídeos ou imagens, por exemplo. Para este estudo de caso, foram selecionados 53579 documentos da base de dominório.

É importante salientar, que exceto o NodeJS, utilizado para a extração dos dados de grupos e páginas do Facebook, este estudo contempla exclusivamente o desenvolvimento de um projeto Java Desktop, versão 8, o qual utiliza API's para manipulação de dados no formato JSON, além do JPA Hibernate na versão 3.0 para persistencia e consultas na base de dados, sendo esta o MySQL versão 5.7.23.

Tabela 5.1: Grupos da rede social Facebook, relacionados às informações da cidade do Rio Grande/R.S.

Identificação do grupo no Facebook	Descrição	Link
RGurgente 220492804809842 261875310896600	Rio Grande Urgente Rio Grande Atento 2 (Oficial) Rio Grande Cidade Nossa (OFICIAL)®	https://www.facebook.com/RGurgente/ https://www.facebook.com/groups/rgatento2oficial/ https://www.facebook.com/groups/261875310896600/
358384060944062 150153225177207 1409957022567659 praiadocassinoicom 1562924313966982 122639974487377 284448841666440 328460373888908 FURGinstitucional2	Prefeitura do Rio Grande O povo do Bolaxa Cassino Beach 2017 Praia do Cassino.com S.O.S Praia Do Cassino Cassino a Maior Praia do Mundo!!! PRAIA DO CASSINO SEST SENAT Rio Grande-RS Universidade Federal do Rio Grande - FURG Rio Grande On Line.	https://www.facebook.com/PrefeituraMunicipaldoRG/ https://www.facebook.com/groups/150153225177207/ https://www.facebook.com/groups/1409957022567659/ https://www.facebook.com/praiadocassinoicom/ https://www.facebook.com/groups/1562924313966982/ https://www.facebook.com/groups/cassinoamaiorpraiadomundo/ https://www.facebook.com/groups/praiadocassino/ https://www.facebook.com/sestsemat.rg/ https://www.facebook.com/FURGinstitucional2
noticiasonline2017 549288371915304 rgatentooficial 763573137050414	Rio Grande Atento só bafão Rio Grande Atento Oficial Rio grande atento Rio Grande - Cidade Histórica Quitanda Cultural Rio Grande Engenharia de Computação - FURG Incubadora Cultura Viva Projeto Lixo Marinho - FURG Rodízio Literário RG Projeto Dunas Costeiras	https://www.facebook.com/noticiasonline2017/ https://www.facebook.com/groups/549288371915304/ https://www.facebook.com/rgatentooficial/ https://www.facebook.com/763573137050414/ https://www.facebook.com/RGHistorica/ https://www.facebook.com/quitandacultural.rg/ https://www.facebook.com/groups/ecompr.furg/ https://www.facebook.com/incubadoraculturaviva https://www.facebook.com/lixomarininhofurg https://www.facebook.com/rodizioliterario https://www.facebook.com/projetodunascosteiras.nema/

Tabela 5.2: Continuação da Tabela 5.1. Grupos da rede social Facebook, relacionados às informações da cidade do Rio Grande/RS.

Identificação do grupo no Facebook	Descrição	Link
smcasrg	SMCAS - Secretaria de Município da Cidadania e Assistência Social Bairro Getúlio Vargas, BGV, Cedro.	https://www.facebook.com/smcasrg/
BairroGetulioVargasBgvCedro 1072810962738904 condominojardimdosol	Bairro Lagoa Rio Grande RS Condomínio Jardim do Sol - Rio Grande	https://www.facebook.com/groups/1072810962738904/ https://www.facebook.com/condominojardimdosol/
357516307667493	BAIRRO SANTA ROSA - RIO GRANDE / RS	https://www.facebook.com/groups/357516307667493/
450103601765861 1432648280394224	Bairro JUNÇÃO - Rio Grande / RS SIRQ - Vila da Quinta: Sociedade Instrução e Recreio da Quinta.	https://www.facebook.com/groups/450103601765861/ https://www.facebook.com/1432648280394224/
ifrsriogrand ccrgsbr smedrg	IFRS - Campus Rio Grande Câmara de Comércio da Cidade do Rio Grande Secretaria de Município da Educação - Rio Grande RS	https://www.facebook.com/ifrsriogrand/ https://www.facebook.com/ccrgsbr/ https://www.facebook.com/smedrg/
museucrg 328100337528866 130323637307218 493351754130735 1472778289626404	Museu da Cidade do Rio Grande Tretas Rio Grande FERG (Futebol em Rio Grande) Guia de Picos de Skate de Rio Grande Diálogos Culturais-SECULT/RG Parque São Pedro Diário da Praça	https://www.facebook.com/museucrg/ https://www.facebook.com/328100337528866/ https://www.facebook.com/groups/130323637307218/ https://www.facebook.com/groups/493351754130735/ https://www.facebook.com/groups/1472778289626404/ https://www.facebook.com/parquesaopedro https://www.facebook.com/diariodapraca

5.2 ANOTAÇÃO DOS DADOS

A anotação dos dados através do sistema web, contou com a participação de 84 diferentes usuários e um total de 1496 documentos anotados. A quantidade de anotações para cada um dos aspectos é apresentada na Tabela 5.3.

Tabela 5.3: Quantidade de documentos anotados em cada aspecto a partir do sistema web.

Aspecto	Quantidade
Economia	157
Educação	318
Empreendedorismo	11
Energia	32
Governança	143
Meio Ambiente	143
Mobilidade	91
Saúde	83
Segurança	455
Tecnologia & Inovação	10
Urbanismo	53

A aquisição e pré-processamento dos dados anotados são descritos na próxima seção deste capítulo.

5.3 AQUISIÇÃO E PRÉ-PROCESSAMENTO DOS DADOS ANOTADOS

De posse do corpus anotado, conforme descrito na metodologia, considerou-se 20% das anotações para tornarem-se documentos de validação, de modo a avaliar a qualidade do classificador. Exceto para os aspectos Economia e Tecnologia & Inovação, que considerou-se 30%, devido ao fato de que estes aspectos contêm poucos documentos anotados. Após a separação dos documentos anotados em treinamento e validação, para os dez experimentos realizados, removeu-se as *stop words* e alguns termos simples. As *stop words* foram anotadas e estão apresentadas na Tabela B.1, localizada no apêndice deste documento. É importante destacar que os termos simples, que não estão contidos na Tabela B.1, são removidos de maneira manual. Estes termos simples referem-se a símbolos, tais como caracteres especiais e de controle, bem como datas e horas.

Nas próximas etapas, ainda no pré-processamento, realizou-se os cálculos de frequência dos termos nos documentos, os quais são necessários para a aplicação das técnicas de mineração EMIM e Phi-Squared, implementadas neste trabalho, bem como a tradução dos termos para utilização em um processo de *Part-Of-Speech Tagger*.

5.4 GERAÇÃO DAS PALAVRAS CANDIDATAS NOS DADOS ANOTADOS

Para esta etapa, após a preparação dos dados, partindo da aquisição e pré-processamento dos documentos anotados, de modo a gerar um conjunto de palavras candidatas para cada aspecto, seleciona-se todos os substantivos anotados nos respectivos corpus de treinamento.

Nesta pesquisa, foram realizados dez experimentos diferentes, com o intuito de obter os melhores termos representativos para cada aspecto. Estes experimentos são discutidos e apresentados na seção referente à geração de termos representativos, ao longo deste capítulo. Porém, em relação às palavras candidatas, a Tabela 5.4 apresenta as quantidades de termos gerados como candidatos para cada aspecto, a partir de cada experimento realizado.

Tabela 5.4: Quantidade de Palavras Candidatas geradas por cada experimento

Aspecto	Experimentos									
	I	II	III	IV	V	VI	VII	VIII	IX	X
Economia	1220	1322	1312	1383	1284	1443	1472	1430	1338	1367
Educação	1955	2079	2210	2236	2303	2202	2198	2269	2228	2247
Empreendedorismo	75	278	272	244	260	260	148	265	140	124
Energia	294	343	347	355	372	363	192	360	324	258
Governança	1755	1888	1860	1936	1868	1995	1902	2021	2031	2044
Meio Ambiente	953	1147	1075	1048	1137	1143	1030	1131	1202	1170
Mobilidade	543	673	682	739	693	721	727	716	684	732
Saúde	868	1161	1115	1069	1076	1022	1062	1098	1106	976
Segurança	1941	2500	2505	2434	2478	2470	2498	2430	2413	2478
Tecnologia & Inovação	183	226	238	130	230	219	140	223	145	243
Urbanismo	487	548	597	554	593	579	583	465	550	554

5.5 DEFINIÇÃO DAS PALAVRAS SEMENTES

Com base na literatura e a partir do conhecimento de domínio sobre cada um dos aspectos, as palavras sementes são anotadas. Estas palavras são fontes de entrada para as técnicas

EMIM e Phi-Squared, as quais são implementadas neste trabalho para a geração dos termos representativos. Nesta seção é importante ressaltar que, o sistema web desenvolvido para anotação dos documentos, também permite que cada publicação possa ser classificada nos seguintes aspectos, diferentes dos mencionados: Alimentação, Cultura, Esporte, Política, Serviço e Trabalho. No entanto, os aspectos Alimentação, Serviço e Trabalho são mesclados na Economia. Já Cultura e Esporte na Educação. Por fim, Política na Governança.

De forma a ilustrar o processo realizado, a Figura 5.5 apresenta as palavras sementes definidas para o aspecto Segurança.

Tabela 5.5: Palavras Sementes para Segurança.

Absolvição	crimes	fuzil	penal
abuso	criminal	guarda	perito
agente	defrec	guardas	pf
apreensão	deic	homicídio	pistola
arma	delação	infracional	poe
assaltado	delagacia	investigação	polícia
assalto	delegado	investigado	policiais
bombeiro	denarc	investigador	preso
bombeiros	denuncia	juizado	processo
brigada	denuncias	justica	reu
busca	desarmamento	lei	revolver
câmera	disparos	leis	roubado
câmeras	dp	maconha	roubo
captura	droga	Mandado	sexual
capturado	drogas	menor	trafico
civil	Embargos	menores	viaturas
cocaina	escrivao	militar	violencia
comarca	estupro	ostensivo	vitima
crack	federal	patrulha	vitimas
crime	foragido	patrulhamento	

As palavras sementes anotadas para os demais aspectos, considerando as mesclas citadas, estão apresentadas a partir das Tabelas C.1, C.2, C.3, C.4, C.5, C.6, C.7, C.8, C.9, C.10 e C.11, todas localizadas no apêndice deste documento.

Com o intuito de avaliar a qualidade das palavras sementes, analisou-se a frequência das mesmas no *dataset* de treinamento. A Tabela 5.6 apresenta para cada um dos experimentos realizados, a fração entre a quantidade de documentos que contêm as palavras sementes anotadas sobre a quantidade de documentos de treinamento.

Tabela 5.6: Experimentos selecionados para cada aspecto

Aspecto	Experimentos									
	I	II	III	IV	V	VI	VII	VIII	IX	X
Economia	85 / 97	103 / 126	107 / 126	107 / 126	108 / 126	108 / 126	106 / 126	105 / 126	107 / 126	3 / 126
Educação	181 / 194	233 / 255	236 / 255	234 / 255	235 / 255	236 / 255	235 / 255	233 / 255	236 / 255	235 / 255
Empreendedorismo	3 / 4	6 / 8	6 / 8	5 / 8	6 / 8	6 / 8	7 / 8	6 / 8	6 / 8	5 / 8
Energia	14 / 16	23 / 26	24 / 26	23 / 26	24 / 26	25 / 26	23 / 26	24 / 26	23 / 26	23 / 26
Governança	80 / 85	108 / 115	106 / 115	103 / 115	103 / 115	106 / 115	105 / 115	104 / 115	103 / 115	103 / 115
Meio Ambiente	69 / 84	94 / 115	96 / 115	100 / 115	96 / 115	100 / 115	99 / 115	94 / 115	98 / 115	94 / 115
Mobilidade	53 / 61	64 / 73	64 / 73	62 / 73	64 / 73	62 / 73	63 / 73	63 / 73	64 / 73	64 / 73
Saúde	46 / 53	55 / 67	55 / 67	54 / 67	56 / 67	58 / 67	53 / 67	56 / 67	55 / 67	54 / 67
Segurança	235 / 260	327 / 364	329 / 364	319 / 364	328 / 364	325 / 364	326 / 364	329 / 364	324 / 364	321 / 364
Tecnologia & Inovação	4 / 5	6 / 7	5 / 7	5 / 7	5 / 7	6 / 7	5 / 7	6 / 7	5 / 7	6 / 7
Urbanismo	35 / 37	39 / 43	40 / 43	40 / 43	39 / 43	40 / 43	41 / 43	41 / 43	42 / 43	40 / 43

5.6 GERAÇÃO DE TERMOS REPRESENTATIVOS E CLASSIFICAÇÃO DOS TEXTOS NOS DADOS ANOTADOS

Para a geração dos termos representativos, foram realizados dez experimentos diferentes aplicando as técnicas EMIM e Phi-Squared, com o intuito de analisar o resultado das métricas de precisão, revocação, f-score e acurácia, de modo a observar a qualidade do classificador. Os dez experimentos estão descritos a partir das Tabelas D.1, D.2, D.3, D.4, D.5, D.6, D.7, D.8, D.9 e D.10, ambas localizadas no apêndice deste documento. Cada experimento classificou os documentos em três abordagens diferentes. I) por termos representativos gerados pelas técnicas propostas, II) considerando somente palavras sementes, e III) através da mescla entre termos representativos e palavras sementes.

É importante citar que após a geração dos termos representativos, realizou-se uma análise manual para definição de quais destes serão realmente utilizados para cada aspecto. Nesta etapa removeu-se alguns termos simples que não foram anotados e também termos sem relevância, tais como horas, datas, símbolos entre outros que foram julgados irrelevantes para o aspecto.

Diante dos resultados das classificações nos dez experimentos diferentes, apresentados nas Tabelas D.1, D.2, D.3, D.4, D.5, D.6, D.7, D.8, D.9 e D.10, localizadas no apêndice deste documento, selecionou-se o melhor classificador para cada aspecto, considerando entre os experimentos a abordagem de termos que obteve maiores valores na métrica F-score. Foi selecionada a métrica F-score justamente por representar a média harmônica entre a precisão e a revocação. As métricas de avaliação para os experimentos e abordagens de termos selecionados para a classificação, são apresentados através da Tabela 5.7, que apresenta, semelhante às

tabelas mencionadas no início deste parágrafo, os dados relativos à matriz de confusão, a qual observa a quantidade de documentos relevantes e irrelevantes recuperados e não recuperados.

Tabela 5.7: Experimentos selecionados para cada aspecto

Aspecto	Experimento	Termos	DR ¹	DNR ²	VP ³	FP ⁴	VN ⁵	FN ⁶	Precisão	Revocação	F-score	Acurácia
Economia	IV	SR ⁷	31	266	22	3	263	9	8.800	7.097	7.857	9.596
Educação	VIII	R ⁸	63	234	53	5	229	10	9.138	8.413	8.760	9.495
Empreendedorismo	III	SR ⁷	3	294	2	0	294	1	10.0	6.667	8.0	9.966
Energia	V	R ⁸	6	291	5	0	291	1	10.0	8.333	9.091	9.966
Governança	VIII	R ⁸	28	269	23	11	258	5	6.765	8.214	7.419	9.461
Meio Ambiente	III	SR ⁷	28	269	24	1	268	4	9.600	8.571	9.057	9.832
Mobilidade	IX	SR ⁷	18	279	12	2	277	6	8.571	6.667	7.500	9.731
Saúde	IV	SR ⁷	16	281	11	0	281	5	10.0	6.875	8.148	9.832
Segurança	IX	R ⁸	91	206	80	7	199	11	9.195	8.791	8.989	9.394
Tecnologia & Inovação	III	S ⁹	3	294	2	1	293	1	6.667	6.667	6.667	9.933
Urbanismo	VII	R ⁸	10	287	6	0	287	4	10.0	6.0	7.5	9.865

¹ Documentos Relevantes

² Documentos Não Relevantes

³ Verdadeiros Positivos

⁴ Falsos Positivos

⁵ Verdadeiros Negativos

⁶ Falsos Negativos

⁷ Sementes e representativas

⁸ Representativas

⁹ Sementes

A partir destes experimentos selecionados, os termos representativos foram gerados e selecionados, e estão ilustrados através das Tabelas E.1, E.2, E.3, E.4, E.5, E.6, E.7, E.8, E.9, E.10, E.11, E.12, E.13, E.14, E.16, E.17, E.18, E.19, E.20 e E.21, todas localizadas no apêndice deste documento.

Considerando que os termos representativos para cada aspecto são obtidos através de experimentos distintos, para analisar a precisão global do classificador, o mesmo foi executado em todo *dataset* anotado. O resultado da precisão global do classificador foi igual a 78%, sendo que nesta análise, verificou-se 1496 documentos, onde obteve 1167 acertos, ou seja, 1167 verdadeiros positivos. Os resultados desta classificação para cada aspecto são ilustrados através da Tabela 5.8.

Com o intuito de apresentar os termos representativos mais frequentes de cada aspecto, nos documentos do *dataset* anotado, gerou-se nuvens de palavras. A Figura 5.1 ilustra uma nuvem de palavras com tais termos para o aspecto Segurança.

Para os demais aspectos, também a partir do *dataset* anotado, os termos representativos

Tabela 5.8: Resultados do classificador final referente à matriz de confusão para os aspectos de estudo deste trabalho.

Aspecto	DR ¹	DNR ²	VP ³	FP ⁴	VN ⁵	FN ⁶	Precisão	Revocação	F-score	Acurácia
Economia	157	1339	136	37	1302	21	7.86	8.66	8.24	9.61
Educação	318	1178	258	24	1154	60	9.15	8.11	8.6	9.44
Empreendedorismo	11	1485	6	4	1481	5	6.0	5.45	5.71	9.94
Energia	32	1464	23	0	1464	9	10.0	7.19	8.36	9.94
Governança	143	1353	119	50	1303	24	7.04	8.32	7.63	9.51
Meio Ambiente	143	1353	117	12	1341	26	9.07	8.18	8.6	9.75
Mobilidade	91	1405	65	30	1375	26	6.84	7.14	6.99	9.63
Saúde	83	1413	57	9	1404	26	8.64	6.87	7.65	9.77
Segurança	455	1041	362	24	1017	93	9.38	7.96	8.61	9.22
Tecnologia & Inovação	10	1486	0	0	1486	10	0	0	0	9.93
Urbanismo	53	1443	24	4	1439	29	8.57	4.53	5.93	9.78

¹ Documentos Relevantes

² Documentos Não Relevantes

³ Verdadeiros Positivos

⁴ Falsos Positivos

⁵ Verdadeiros Negativos

⁶ Falsos Negativos

mais frequentes são ilustrados através das Figuras F.1, F.2, F.3, F.4, F.5, F.6, F.7, F.8, F.9 e F.10, ambas localizadas no apêndice deste documento. É importante destacar que o tamanho de cada palavra está relacionado à frequência de ocorrência desta nos documentos anotados. Além disso, não são ilustrados todos os termos representativos, somente os mais frequentes.

5.7 CLASSIFICAÇÃO E GERAÇÃO DOS TERMOS REPRESENTATIVOS NOS DADOS NÃO ANOTADOS

A partir dos termos representativos gerados e selecionados na etapa anterior, realizou-se a classificação em todo conjunto de documentos da base de dados. Os resultados desta classificação considerando 53579 documentos são ilustrados a partir da Tabela 5.9.

Diante dos documentos classificados, conforme os resultados apresentados na Tabela 5.9, aplicou-se novamente as técnicas EMIM e Phi-Squared para geração de novos termos representativos de cada aspecto. Nesta abordagem, foram considerados todos os documentos classificados no respectivo aspecto. Estes termos estão apresentados através das Tabelas G.1, G.2, G.3, G.4, G.5, G.6, G.7, G.8, G.9, G.10, G.11, G.12, G.13, e G.14, localizadas no apêndice deste documento. É importante salientar que para o aspecto Tecnologia & Inovação não foram



Figura 5.1: Nuvem de palavras para o aspecto Segurança.

Tabela 5.9: Quantidade de documentos classificados por aspecto, considerando os respectivos termos representativos gerados e selecionados, apresentados nas Tabelas E.1, E.2, E.3, E.4, E.5, E.6, E.7, E.8, E.9, E.10, E.11, E.12, E.13, E.14, E.16, E.17, E.18, E.19, E.20 e E.21, localizadas no apêndice deste documento.

Aspecto	Quantidade
Economia	8462
Educação	9098
Empreendedorismo	341
Energia	705
Governança	6217
Meio Ambiente	2577
Mobilidade	2754
Saúde	1710
Segurança	9726
Tecnologia & Inovação	78
Urbanismo	394
Não classificados	11517
Total de documentos	53579

gerados termos representativos, devido ao fato de não ter encontrado palavra(s) semente(s) nos documentos classificados.

A Figura 5.2 ilustra a nuvem de palavras com os termos representativos mais frequentes para o aspecto Segurança.



Figura 5.2: Nuvem de palavras para o aspecto Segurança.

Para os demais aspectos, os termos representativos mais frequentes estão apresentados através das Figuras H.1, H.2, H.3, H.4, H.5, H.6, H.7, H.8 e H.9, localizadas no apêndice deste trabalho.

Percebe-se que os termos representativos mais frequentes gerados a partir *dataset* anotado são diferentes dos obtidos através da classificação de todos documentos. Entretanto, observa-se que os métodos aplicados em todo corpus classificado, geram indícios importantes não identificados na base de dados anotados, mas que emergem da mineração.

6 CONSIDERAÇÕES FINAIS

Este trabalho apresentou como objetivo estudar e propor uma metodologia baseada em classificação de textos para ser aplicada a postagens em redes sociais, com o intuito de obter indícios sobre a percepção coletiva da população a partir de termos representativos relacionados à indicadores de qualidade de vida no âmbito de cidades inteligentes. A proposta foi validada em um estudo de caso associado a dados textuais extraídos de páginas e grupos temáticos do Facebook, relacionados às informações da cidade do Rio Grande / RS.

O trabalho envolveu uma revisão bibliográfica associada aos conceitos referentes a indicadores para cidades inteligentes. Também foram estudadas técnicas de mineração de dados não estruturados e suas métricas de avaliação.

Além de realizar um estudo bibliográfico, os objetivos desta pesquisa foram alcançados por meio de um modelo computacional de extração, pré-processamento, mineração e pós-processamento dos dados textuais. Este modelo propôs o uso de técnicas de classificação baseadas em termos representativos para que a partir destes possam ser extraídos mapas conceituais sobre os principais indicadores que compoem o nível de inteligência de um município. Ademais, essa investigação assumiu o desafio tecnológico de realizar a classificação em textos de publicações fracamente estruturadas, considerando que os dados analisados contêm pouca ou quase nenhuma estrutura.

O modelo foi aplicado a um estudo de caso associado a obtenção de termos representativos e classificação de postagens do Facebook no contexto da cidade do Rio Grande. Foi gerado um *dataset* anotado com 1496 postagens, a qual foi utilizada para validação da proposta. Vale destacar que a metodologia apresentada nesta dissertação também pode ser aplicada em dados provindos de outras redes sociais.

Os resultados obtidos apresentam indícios que o método pode ser aplicado, uma vez que as métricas de avaliação foram adequadas. Bons resultados foram obtidos sobretudo para os aspectos com maior número de dados anotados, o que indica que esforços terão que ser realizados para a obtenção de uma maior quantidade de amostras para treinamento. Inclusive, esta expansão do *dataset* já está sendo realizada. Ademais, os termos representativos e as classificações obtidas neste estudo, podem ser utilizados em etapa de pós-processamento para a obtenção

de um discurso coletivo sobre a percepção do cidadão em relação ao nível de inteligencia da sua cidade.

Como trabalhos futuros pretende-se realizar mais testes bem como uma análise mais detalhada dos dados, seja em termos quantitativos (a partir do *dataset* expandido) como qualitativos tentando explicar as métricas obtidas.

A aplicação de técnicas para mineração de opinião a partir dos resultados obtidos serão trabalhos futuros.

REFERÊNCIAS

- AGGARWAL, C. C.; ZHAI, C. *Mining text data*. [S.l.]: Springer Science & Business Media, 2012. Citado na página 47.
- ANDRADE, D. C. de. Priorização de nós de uma rede de sensores sem fio com base na análise de dados de redes sociais. *Anais Seminário de Iniciação Científica*, n. 21, 2017. Citado 4 vezes nas páginas 17, 58, 59 e 60.
- BARALIS, E.; CAGLIERO, L.; MAHOTO, N.; FIORI, A. Graphsum: Discovering correlations among multiple terms for graph-based summarization. *Information Sciences*, Elsevier, v. 249, p. 96–109, 2013. Citado na página 47.
- BERRONE, P.; RICART, J. E.; CARRASO, C.; RICART, R. *Iese: Cities in motion index 2016. IESE, University of Navarra Business School, New York, USA*, 2016. Citado na página 34.
- CABENA, P.; HADJINIAN, P.; STADLER, R.; VERHEES, J.; ZANASI, A.; JOSE, C. I. B. M. C. S.; JOSE, C. I. T. S. O. S. *Discovering data mining: from concept to implementation*. [S.l.]: Prentice Hall PTR New Jersey, 1997. Citado na página 27.
- CARAGLIU, A.; BO, C. D.; NIJKAMP, P. Smart cities in europe. *Journal of urban technology*, Taylor & Francis, v. 18, n. 2, p. 65–82, 2011. Citado na página 33.
- CHURCH, K. W.; GALE, W. A. Concordances for parallel text. In: *Proceedings of the Seventh Annual Conference of the UW Centre for the New OED and Text Research*. [S.l.: s.n.], 1991. p. 40–62. Citado 2 vezes nas páginas 48 e 49.
- CHURCH, K. W.; HANKS, P. Word association norms, mutual information, and lexicography. *Computational linguistics*, MIT Press, v. 16, n. 1, p. 22–29, 1990. Citado na página 48.
- CHURCH, K. W.; MERCER, R. L. Introduction to the special issue on computational linguistics using large corpora. *Computational linguistics*, MIT Press, v. 19, n. 1, p. 1–24, 1993. Citado na página 48.
- CORDEIRO, A. D. et al. Gerador inteligente de sistemas com auto-aprendizagem para gestão de informações e conhecimento. Florianópolis, SC, 2005. Citado na página 41.
- DIAS, L. de C. Análise de redes sociais com uso de aprendizado de máquina para prever o tráfego de veículos em zonas urbanas. 2017. Citado 6 vezes nas páginas 17, 53, 54, 55, 56 e 57.
- EBECKEN, N. F.; LOPES, M. C. S.; COSTA, M. C. et al. Mineração de textos. *Sistemas inteligentes: fundamentos e aplicações*. São Carlos: Manole, p. 337–370, 2003. Citado na página 43.
- FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. *AI magazine*, v. 17, n. 3, p. 37, 1996. Citado na página 38.
- FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. The kdd process for extracting useful knowledge from volumes of data. *Communications of the ACM*, ACM, v. 39, n. 11, p. 27–34, 1996. Citado na página 39.

FIGUEREDO, M.; SOUSA, A. Utilizando redes sociais para detectar crimes em tempo real. 2016. Citado 2 vezes nas páginas 17 e 52.

GIFFINGER, R.; GUDRUN, H. Smart cities ranking: an effective instrument for the positioning of the cities? *ACE: architecture, city and environment*, Centre de Política del Sòl i Valoracions-Universitat Politècnica de Catalunya, v. 4, n. 12, p. 7–26, 2010. Citado na página 33.

HALL, R. E.; BOWERMAN, B.; BRAVERMAN, J.; TAYLOR, J.; TODOSOW, H.; WIMMERSPERG, U. V. *The vision of a smart city*. [S.l.], 2000. Citado 2 vezes nas páginas 28 e 33.

HAN, J.; PEI, J.; KAMBER, M. *Data mining: concepts and techniques*. [S.l.]: Elsevier, 2011. Citado 2 vezes nas páginas 17 e 36.

HARRISON, C.; DONNELLY, I. A. A theory of smart cities. In: *Proceedings of the 55th Annual Meeting of the ISSS-2011, Hull, UK*. [S.l.: s.n.], 2011. v. 55, n. 1. Citado na página 33.

JUNIOR, J. R. C. Desenvolvimento de uma metodologia para mineração de textos. *Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro*, 2007. Citado 2 vezes nas páginas 41 e 43.

KHATOUN, R.; ZEADALLY, S. Smart cities: concepts, architectures, research opportunities. *Communications of the ACM*, ACM, v. 59, n. 8, p. 46–57, 2016. Citado na página 34.

KOMNINOS, N. *Intelligent cities: innovation, knowledge systems and digital spaces*. [S.l.]: Routledge, 2013. Citado na página 33.

LEVENSHTEIN, V. I. Binary codes capable of correcting deletions, insertions, and reversals. In: *Soviet physics doklady*. [S.l.: s.n.], 1966. v. 10, n. 8, p. 707–710. Citado na página 54.

LOPES, M. C. S. Mineração de dados textuais utilizando técnicas de clustering para o idioma português. *Rio de Janeiro*: sn, 2004. Citado na página 40.

MADEIRA, R. d. O. C. *Aplicação de técnicas de mineração de texto na detecção de discrepâncias em documentos fiscais*. Tese (Doutorado), 2015. Citado na página 37.

MANNING, C. D.; MANNING, C. D.; SCHÜTZE, H. *Foundations of statistical natural language processing*. [S.l.]: MIT press, 1999. Citado na página 48.

MORAIS, E. A. M.; AMBRÓSIO, A. P. L. Mineração de textos. *Relatório Técnico-Instituto de Informática (UFG)*, 2007. Citado 8 vezes nas páginas 17, 37, 41, 42, 43, 44, 45 e 46.

MOURA, M. Proposta de utilização de mineração de textos para seleção, classificação e qualificação de documentos. *Embrapa Informática Agropecuária-Documentos (INFOTECA-E)*, Campinas: Embrapa Informática Agropecuária, 2004., 2004. Citado na página 40.

PALAZZO, M.; LOH, S.; AMARAL, L.; WIVES, L. Descoberta de conhecimento em textos através da análise de seqüências temporais. In: *Workshop em Algoritmos e Aplicações de Mineração de Dados-WAAMD, SBBD: Sociedade Brasileira de Computação*. [S.l.: s.n.]. v. 2, p. 49–56. Citado na página 39.

PASSOS, E.; GOLDSCHMIDT, R. Data mining: um guia prático. *Editora Campus, Rio de Janeiro*, 2005. Citado 2 vezes nas páginas 17 e 38.

RAMOS, H. d. S. C.; BRÄSCHER, M. Aplicação da descoberta de conhecimento em textos para apoio à construção de indicadores infométricos para a área de c&t. *Ciência da Informação*, v. 38, n. 2, 2009. Citado na página 27.

SALTON, G.; MCGILL, M. J. Introduction to modern information retrieval. McGraw-Hill, Inc., 1986. Citado na página 41.

SÁPIRAS, L. A. Mineração de opiniões em aspectos em fontes de opiniões fracamente estruturadas. 2015. Citado na página 48.

SCHAFFER, C. L. Monitoramento de mídias sociais no processo de análise da percepção dos usuários sobre a segurança de porto alegre. 2016. Citado na página 51.

SÉRGIO, M. C.; SILVA, T. do Nascimento da; GONÇALVES, A. L. Descoberta de conhecimento a partir de informações não estruturadas por meio de técnicas de correlação e associação. *Em Questão*, Universidade Federal do Rio Grande do Sul, v. 22, n. 2, 2016. Citado na página 47.

STATISTA, A. The statistics portal. *Web site*: <https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/>, 2017. Citado 2 vezes nas páginas 17 e 30.

STATISTA, A. The statistics portal. *Web site*: <https://www.statista.com/statistics/244936/number-of-facebook-users-in-brazil/>, 2017. Citado 3 vezes nas páginas 17, 30 e 62.

STATISTA, A. The statistics portal. *Web site*: <https://www.statista.com/statistics/254734/most-popular-social-networking-sites-in-brazil/>, 2017. Citado 2 vezes nas páginas 17 e 31.

SYSTEM, U. *Ranking Connected Smart Cities*. 2017. [Accessado 31-08-2018]. Disponível em: <<http://www.connectedsmartcities.com.br/resultados-do-ranking-connected-smart-cities/?lang=en>>. Citado 2 vezes nas páginas 28 e 34.

SYSTEMS, A. U. Inteligencia de mercado | urban systems brasil. *Web site*: <https://www.urbansystems.com.br>, 2018. Citado na página 34.

TAN, P.-N.; STEINBACH, M.; KUMAR, V. *Introduction to data mining*. 1st. [S.l.]: Boston: Pearson Addison Wesley. xxi, 2005. Citado 2 vezes nas páginas 37 e 47.

VELOSO, R. A. Geopantin: Uma ferramenta de rede social baseada em localização inspirada em soluções de cidades inteligentes. 2012. Citado na página 57.

WEISS, M. C. Cidades inteligentes como nova prática para o gerenciamento dos serviços e infraestruturas urbanas: estudo de caso da cidade de porto alegre/marcos cesar weissl; orientação: Profº drº roberto carlos bernardes. 2013. Citado na página 28.

WIVES, L. Recursos de text mining. *Disponível por www em http://www.inf.ufrgs.br/~wives/portugues/textmining.html (15 de maio de 2001)*, 2005. Citado na página 39.

WIVES, L. K. Tecnologias de descoberta de conhecimento em textos aplicadas à inteligência competitiva. *Exame de Qualificação EQ-069, PPGC-UFRGS*, 2002. Citado 4 vezes nas páginas 37, 41, 44 e 46.

ZHANG, C.; ZHOU, G.; YUAN, Q.; ZHUANG, H.; ZHENG, Y.; KAPLAN, L.; WANG, S.; HAN, J. Geoburst: Real-time local event detection in geo-tagged tweet streams. In: ACM. *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. [S.l.], 2016. p. 513–522. Citado na página 59.

APÊNDICE A – TELAS DO SISTEMA WEB PROPOSTO PARA ANOTAÇÃO DOS DOCUMENTOS

The screenshot shows a user creation form. At the top, there is a label "Username" above a text input field containing the placeholder "Enter Username". Below it is a label "Password" above another text input field containing the placeholder "Enter Password". At the bottom of the form is a large purple rectangular button with the word "Create" in white.

Figura A.1: Tela de criação de usuário do sistema web proposto.

The screenshot shows a login form. At the top, there is a label "Username" above a text input field containing the placeholder "Enter Username". Below it is a label "Password" above another text input field containing the placeholder "Enter Password". At the bottom of the form is a large purple rectangular button with the word "Login" in white.

Figura A.2: Tela de login do sistema web proposto.

[Sair](#)

Classificação de documentos do Facebook

[CLASSIFICADOS](#) [NÃO CLASSIFICADOS](#)

ID:129856

DATA:14/02/2017

NAME:
Rio Grande On Line.

DESCRIPTION:
Rio Grande On LineRegistrados dois acidentes com ciclistas nas últimas horas.....

Acidente na ERS 734 próximo ao shopping
Uma ciclista de 33 anos foi atropelada por um carro na Ers 734 na tarde desta Segunda Feira,foi socorrido e encaminhado a santa casa.

Acidente na Avenida no Cassino
Uma ciclista de 20 anos foi atropelada por volta das 21:00 hs na Avenida Rio Grande no Cassino por um carro. A vitima teve diversos ferimentos e foi encaminhada a santa casa de Rio Grande.

Fonte : Brigada Militar
Foto : Ilustrativa
Texto e Postagem : Fabiano Correa

MESSAGE:

Alimentação
Cultura
Economia
Educação
Empreendedorismo

Figura A.3: Tela de apresentação das publicações no sistema web proposto. Parte 1.

Alimentação
Cultura
Economia
Educação
Empreendedorismo
Energia
Esporte
Governança (Administração pública)
Meio Ambiente
Mobilidade
Política
Saúde
Segurança
Serviço
Tecnologia e Inovação
Trabalho
Urbanismo

De acordo com a sua opinião, como você define este texto em relação ao conteúdo da publicação? O texto expressa um sentimento:

- Negativo
 Neutro
 Positivo

[Enviar / Próximo](#)



Universidade Federal do Rio Grande - FURG
Programa de Pós-Graduação em Modelagem Computacional

Figura A.4: Tela de apresentação das publicações no sistema web proposto. Parte 2.

ID:129856

DATA:14/02/2017

NAME:
Rio Grande On Line.

DESCRIPTION:
Rio Grande On Line Registrados dois acidentes com ciclistas nas últimas horas.....

Acidente na ERS 734 próximo ao shopping
Uma ciclista de 33 anos foi atropelado por um carro na Ers 734 na tarde desta Segunda Feira, foi socorrido e encaminhado a santa casa.

Acidente na Avenida no Cassino
Uma ciclista de 20 anos foi atropelada por volta das 21:00 hs na Avenida Rio Grande no Cassino por um carro. A vitima teve diversos ferimentos e foi encaminhada a santa casa de Rio Grande.

Fonte : Brigada Militar
Foto : Ilustrativa
Texto e Postagem : Fabiano Correa

MESSAGE:

Alimentação
 Cultura
 Economia
 Educação
 Empreendedorismo
 Energia
 Esporte
 Governança (Administração pública)
 Meio Ambiente
 Mobilidade
 Política
 Saúde
 Segurança
 Serviço
 Tecnologia e Inovação
 Trabalho
 Urbanismo

De acordo com a sua opinião, como você define este texto em relação ao conteúdo da publicação? O texto expressa um sentimento:

Negativo
 Neutro
 Positivo

[Enviar / Próximo](#)

Figura A.5: Tela de apresentação das publicações no sistema web proposto. Parte 3.

```

SELECT Document.id,
       Document.name,
       Document.description,
       Document.message,
       Substr(Document.createdtime, 1, 10) AS createdTime
  FROM documenti AS Document
 LEFT JOIN documentclassification
        ON documentclassification.iddocumento = Document.id
 WHERE ( NOT Isnull(Document.message)
        OR NOT Isnull(Document.description) )
        AND Document.ehRioGrande = 1 AND Document.id NOT IN
        ( SELECT DocumentClassification.idDocumento
          FROM DocumentClassification GROUP BY DocumentClassification.idDocumento
          HAVING COUNT(DocumentClassification.idDocumento) >= 5 ) AND
        Document.id NOT IN ( SELECT DocumentClassification.idDocumento
                                FROM DocumentClassification WHERE ISNULL(DocumentClassification.alimentacao)
                                AND ISNULL(DocumentClassification.cultura) AND ISNULL(DocumentClassification.economia)
                                AND ISNULL(DocumentClassification.educacao) AND ISNULL(DocumentClassification.empreendedorismo)
                                AND ISNULL(DocumentClassification.energia) AND ISNULL(DocumentClassification.esporte)
                                AND ISNULL(DocumentClassification.governanca) AND ISNULL(DocumentClassification.meioAmbiente)
                                AND ISNULL(DocumentClassification.mobilidade) AND ISNULL(DocumentClassification.politica)
                                AND ISNULL(DocumentClassification.saude) AND ISNULL(DocumentClassification.seguranca)
                                AND ISNULL(DocumentClassification.servico) AND ISNULL(DocumentClassification.tecnologiaInovacao)
                                AND ISNULL(DocumentClassification.trabalho) AND ISNULL(DocumentClassification.urbanismo) )
                                AND Document.id NOT IN ( SELECT DocumentClassification.idDocumento
                                FROM DocumentClassification WHERE DocumentClassification.idUserClassificationWebSystem = 1 )
        GROUP BY Document.id, Document.name, Document.description, Document.message ORDER BY COUNT(Document.id) DESC,
        Document.id IN ( SELECT DocumentClassification.idDocumento FROM DocumentClassification WHERE (
        NOT ISNULL(DocumentClassification.alimentacao) OR NOT ISNULL(DocumentClassification.cultura) OR
        NOT ISNULL(DocumentClassification.economia) OR NOT ISNULL(DocumentClassification.educacao) OR
        NOT ISNULL(DocumentClassification.empreendedorismo) OR NOT ISNULL(DocumentClassification.energia) OR
        NOT ISNULL(DocumentClassification.esporte) OR NOT ISNULL(DocumentClassification.governanca) OR
        NOT ISNULL(DocumentClassification.meioAmbiente) OR NOT ISNULL(DocumentClassification.mobilidade) OR
        NOT ISNULL(DocumentClassification.politica) OR NOT ISNULL(DocumentClassification.saude) OR
        NOT ISNULL(DocumentClassification.seguranca) OR NOT ISNULL(DocumentClassification.servico) OR
        NOT ISNULL(DocumentClassification.tecnologiaInovacao) OR NOT ISNULL(DocumentClassification.trabalho) OR
        NOT ISNULL(DocumentClassification.urbanismo) ) GROUP BY DocumentClassification.id
        ORDER BY COUNT(DocumentClassification.idDocumento) ) DESC, RAND() LIMIT 1 OFFSET 0
  
```

Figura A.6: Query para apresentação dos documentos no sistema web proposto.

```

CREATE TABLE WordCloudTF (
    id bigint(20) NOT NULL AUTO_INCREMENT,
    classeGramatical varchar(255) DEFAULT NULL,
    palavra varchar(255) DEFAULT NULL,
    palavraENG varchar(255) DEFAULT NULL,
    quantidade bigint(20) DEFAULT NULL,
    idDocumento bigint(20) DEFAULT NULL,
    PRIMARY KEY (id),
    KEY FKD52AFF85F7177E59 (idDocumento),
    CONSTRAINT FKD52AFF85F7177E59
        FOREIGN KEY (idDocumento)
        REFERENCES Document (id)
) ENGINE=InnoDB AUTO_INCREMENT=1
    DEFAULT CHARSET=latin1

```

Figura A.7: Query SQL para a criação da tabela *WordCloudTF*.

```

INSERT INTO WordCloudTF(palavra, quantidade, idDocumento)
SELECT WordCloud.palavra,
       IF(COUNT(WordCloud.palavra) IS NULL, 0,
          COUNT(WordCloud.palavra) ) AS quantidade,
       WordCloud.idDocumento
FROM WordCloud
WHERE LENGTH(WordCloud.palavra) > 3
GROUP BY WordCloud.palavra, WordCloud.idDocumento
ORDER BY WordCloud.palavra, WordCloud.idDocumento ASC;

```

Figura A.8: Query SQL para a população da tabela *WordCloudTF* a partir dos registros da tabela *WordCloud*.

APÊNDICE B – *STOP WORDS* E TERMOS SIMPLES ANOTADOS

Tabela B.1: *Stop Words* e alguns termos simples anotados.

!	antes	de	dois	estive	fôramos	houvéssemos	no	porque	setor	terão	três	—
!!	ao	dela	dos	estiver	forem	início	nome	posso	seu	terei	tu	}
!!!	aos	delas	duas	estivera	forma	isto	nos	pouca	seus	teremos	tua	
#	após	dele	e	estivéramos	formos	nossa	poucas	si		teria		tuas
%	aquela	deles	ela	estiverem	fosse	já	nossas	pouco	sido	teriam	tudo	
&	aquelass	depois	elas	estivermos	fossem	janeiro	nossa	poucos	silva	teríamos	um	
,	aquele	desde	ele	estivéssemos	fôssemos	la	nossos	pt	só	teu	uma	
(aqueles	dessa	elos	estou	fui	lado	nova	quais	sob	teus	umas	
)	aqui	dessas	em	eu	há	lei	novo	qual	sobre	teve	uns	
*	aquilo	desse	então	eu	hajamos	lhe	num	que	somos	ti	us	
+	área	desses	entre	exemplo	hão	lhes	numa	quem	sou	tido	vai	
-	as	desta	era	falta	havemos	lo	O	quer	sp	tinha	vão	
.	assim	destas	eram	fato	hei	local	onde	são	sua	tinham	ven	
/	até	deste	éramos	faz	hoje	lugar	os	se	suas	tínhamos	vendo	
:	através	destes	especial	fazendo	houve	mas	ou	segundo	su	tive	ver	
;	bem	deve	essa	fazer	houvemos	me	oura	seja	talvez	tivemos	vez	
<	bom	dever	essas	feita	houver	mês	outras	sejam	também	tiver	vezes	
=	cada	deverá	esse	feitas	houvera	mil	outro	sejamos	tampouco	tivera	vida	
>	casa	deveria	esses	fez	houveram	na	outros	sem	tão	tiveram	vindo	
?	cerca	deveriam	esta	fhc	houvéramos	nada	para	semana	te	tivéramos	vir	
à	coisa	devia	estado	filho	houverão	não	parte	sempre	tel	tiverem	você	
ainda	coisas	deviam	estados	filme	houverei	nas	partir	sendo	tem	tivermos	vocês	
além	com	dia	estão	fim	houverem	nem	paulo	ser	temos	tivesse	vos	
algum	como	disse	estas	final	houveremos	nenhum	pela	será	tendo	tivessem	[
algum	contudo	disso	estávamos	foi	houveria	nessa	pelas	serão	tenha	tivessemos	\	
alguma	da	disto	este	folha	houveriam	nessas	pelo	serei	tenham	toda]	
alguns	daquele	dito	esteja	fomos	houveríamos	nesta	pelos	seremos	tenhamos	todas	^	
ano	daqueles	diz	estejam	for	houvermos	nestas	per	seria	tenho	todavia	—	
anos	dar	dizem	estejamos	fora	houvesse	neste	por	seriam	ter	todo	—	
ante	das	do	estes	foram	houvessem	ninguém	potém	seríamos	terá	todos	{	

APÊNDICE C – PALAVRAS SEMENTES ANOTADAS PARA CADA ASPECTO

Tabela C.1: Palavras Sementes para Economia.

alimentação	contratação	Governo	paralisação	Salário
Ambiente	currículo	greve	Patrão	SEBRAE
APL	demissão	hierarquia	pesca	Segurança
aplicação	Desemprego	Imposto	Pessimista	Seguridade
arranjo	dinheiro	Indivíduo	PIB	seguro
atendimento	dívida	Indústria	pizza	serviço
automação	Economia	Inovação	Política	serviços
autônomo	empatia	investimento	Políticas	Servidor
banco	empreendedor	Jornada	pólo	SESI
bauru	empreendimento	julgamento	ponto	Shopping
bebida	empregado	lanche	Porto	sindicância
burocracia	empregador	lancheria	portos	sindicato
camarão	emprego	leilão	prestaçao	sistema
capacitação	empresa	Local	produção	social
cartão	empresas	logística	produtivo	soja
cdl	entrega	loja	professor	taxa
chamada	erro	Manifestação	Profissional	tecnologias
chefia	estabelecimento	Maquinário	Programa	tecon
Cliente	feira	metalúrgico	Protesto	tele-entrega
Colaboração	fertilizante	milho	pública	terminal
comerciante	fertilizantes	moradia	qualificação	trabalho
comerciário	Fila	naval	quentinha	treinamento
comércio	fome	negócio	Recursos	Valor
comércios	Força	obra	relacionamento	venda
comida	formação	oferta	renda	vende
Compra	fornecimento	oportunidade	restaurante	vendedor
Confiança	funcionário	organizacional	RH	Xis
container	gerente	pagamento	roubo	

Tabela C.2: Palavras Sementes para Educação (parte 01).

academia	capoeira	enem	hipismo	modernismo	portuguesa
africana	carnaval	ensino	história	moderno	prancha
afrobrasileira	cartada	equipe	hobby	motociclismo	práxis
agachamento	cartismo	equipes	identidade	mudança	premiação
agradece	categorias	escanteio	iemanjá	multicultural	prêmio
aluno	ccmar	escola	igreja	municipais	prenda
alunos	cesta	escolar	imagem	municipal	prendadas
ambiental	cidade	escolas	imagens	musculação	processo
amigo	cinema	escultura	incultos	museu	procissão
antidesportista	civilização	espaço	indígena	música	professor
antropologia	clássica	especialização	instrução	nadadores	professores
apito	cocheira	espetáculo	integração	nado	profissional
apresentação	competição	esportividade	intelectual	natação	programa
aquático	comunicação	estádio	jiu-jitsu	obediencia	projeto
árbitro	contemplação	estátua	jogando	objetos	prova
arqueologia	contemporâneo	etnia	jogo	octógono	pscina
artefato	correr	etnografia	jogos	olimpíada	quadra
artes	correu	evento	juiz	olimpíadas	raquete
artesanato	corrida	exposição	karatê	onda	recreação
ataque	corridas	falta	kart	oral	rede
atividade	creche	fantasia	Kartismo	paidéia	regata
atleta	cultura	farroupilha	kartista	paleoíndio	reintegração
atlético	cultural	feira	kitesurf	papo-cabeça	religião
atletismo	culturalmente	festa	kitesurfe	paraolimpíada	remo
ator	cumprimento	festival	lançamento	paraolimpíadas	renascimento
autógrafos	curso	fitness	letramento	passatempo	respeitar
automobilismo	cursos	folclore	líbero	patinação	ringue
automobilística	dança	folclorização	libras	pattins	roller
avaliar	desatenção	foto	licença	peão	rua
baile	desenho	fotografia	linguas	pedagógico	rústica
banda	desfile	futebol	livraria	penalt	sacadas
bandeirinha	desobediente	futilidade	livro	penalti	sacro
barco	desportista	gaúcho	ludicidade	percepção	secult
baseball	didatismo	ginásio	lúdico	periferia	sedentária
basquete	digital	gincana	luta	pescaria	sedentário
batom	disputa	globalização	má-educação	pintura	sesc
berimbau	dissociação	gol	manifestação	pista	show
bicicleta	divertimento	gótilo	manual	pluridimensional	símbolo
bóia	doutorado	gráfico	maratona	pobre	sincretismo
bola	doutrina	grafite	mec	pódium	sítio
brasileiro	eclética	grauduação	medalha	poetisa	skate
brincadeira	educação	halterofilismo	merenda	pólo	slack
campeonato	empenhada	handball	mestrado	pontuação	smed
canção	empenhado	hapkido	moda	popular	sociedade
cancha	empenho	hegemônica	modalidade	popularmente	sócioeducativo

Tabela C.3: Palavras Sementes para Educação (parte 02).

sonata	tempo	urbano
standup	tenis	vela
subcultura	tennis	vento
superior	time	vídeo
surf	torneio	visual
surfe	tradicionalismo	vitória
taekwondo	travessia	volei
tapetão	triathlon	windsurf
teatro	trinador	windsurfe
telento	troféu	zumba
templo	turfe	

Tabela C.4: Palavras Sementes para Empreendedorismo.

Aceleradora	fearg	mar
Afiliados	fecis	Marketing
Aporte	feira	Mercado
B2B	feiras	Métrica
B2C	fejunca	Métricas
Bootstrapping	festa	negócio
Break-even	filial	negócios
Budget	Follow-up	Networking
Business	franquia	orçamento
Capital	franquias	Outsourcing
Co-investimento	giro	Patente
corporativo	guia	planejamento
Coworking	incubadora	recursos
empreendedorismo	Intraempreendedorismo	Remarketing
empreendimento	Investidor-anjo	Stakeholders
empresa	LMS	Startup
empresas	lucro	Startups
Escalabilidade	lucros	turístico

Tabela C.5: Palavras Sementes para Energia.

Abeer	eletrecidade	poste
Aben	Eletrobrás	potencia
Abrace	Eletrointensivos	Racionalização
Abraceel	energia	racionamento
Abradee	eólica	Renováveis
aérogeradores	fotovoltaica	Reservatórios
ANEEL	gás	Sazonalização
Anti-apagão	gato	sistema
apagão	geração	Solar
Biomassa	gerador	Subestações
CCEE	hidrelétrica	tensao
ceee	hidrelétricas	Termeletricidade
Co-geração	iluminação	térmica
combustão	lâmpada	Topologia
consumo	lâmpadas	Transmissão
Curva-guia	luz	usina
Desligamento	modulacao	volt
Desligamentos	nuclear	voltagem
distribuicao	ONS	watt

Tabela C.6: Palavras Sementes para Governança.

acobertado	corrupção	gestão	necessidades	público
ações	corrupto	Governabilidade	Objetivo	recuperação
administração	criação	Governança	oficial	Recurso
Administrador	cuidadoso	governar	Ofício	Roubo
ajuizado	CUT	Governo	Oneroso	sábio
Ambiente	Dados	greve	Orçamento	secretaria
apurado	Decisão	humanitário	organização	Secretarias
Argumento	democracia	Ignorado	pagamento	Secreto
Atento	deputado	Ilegítimo	parlamentar	senador
ativista	descrição	ilusório	partido	Sessão
Autoridade	desvio	imposto	Pessimista	Sincoergs
baldio	diário	Incompetência	Petrobrás	SINDIAGRO
Balizado	diplomacia	incompetente	PIB	sindicato
Banco de dados	diplomata	Informações	Planejamento	SINTERG
barato	diplomático	inteligente	polícia	Sirecom Sul
câmara	discreto	Interesses	Política	sistema
campanha	efetividade	internacional	Políticas	Sistêmico
Caro	Eficiência	investimento	ponderado	social
Chefia	eleições	juiz	população	socialização
Cidade	eleito	juíza	Posição	solução
Cível	enganados	jurista	povo	soluções
Civilidade	erróneo	ladrão	Prefeito	STF
clandestino	Escândalo	lava-jato	Prefeitura	subentendido
coletivo	escrupuloso	lei	Presidência	Supremo
coligação	Estado	licitação	presidenta	Tecnologia
competente	estratégias	Máfia	presidente	traídos
Comunicação	evento	mafioso	princípios	Transparência
comunidade	falso	manifestação	prisão	tribunal
Conceitos	farsante	metas	processos	urbanidade
Condescendente	Fechamento	ministério	Programas	Valor
conhecimento	federal	ministra	projeto	vereador
contexto	Ferramentas	ministro	Projetos	vereadores
Conversa	flexível	municipal	pública	visão
correto	gerenciamento	Município	públicas	Votação

Tabela C.7: Palavras Sementes para Meio Ambiente.

abate	Clorofluorcarbono	Impacto
agrotoxico	CO2	Impactos
agrotoxicos	Combustíveis	lixo
alcaloide	danos	Mitigação
alcaloides	decomposicao	nublado
ambiental	Decompositores	organico
ambiente	defeso	organicos
biodegradavel	Desmatamento	pesca
Biodiesel	Dióxido	Poluente
Biodiversidade	Ecodesign	poluição
Biogás	Ecoeficiência	preservacao
calor	ecologia	queimada
camarão	ecológico	queimadas
Carbonatos	ecológicos	residuos
Carbono	Efluentes	sol
chuva	fogo	sonora
chuvas	fósseis	sustentabilidade
clima	frio	Sustentável
climática	geada	temperatura
climático	granizo	temperaturas
climáticos	Habitat	Transgênico

Tabela C.8: Palavras Sementes para Mobilidade.

acessibilidade	ERS	rotatória
acesso	estacionamento	rotatórias
acidente	etanol	ruas
acidentes	fluxo	semáforo
avenida	gasolina	semáforos
avenidas	guarda	sinalização
bicicleta	iluminação	SMTT
bloqueado	infração	subsídio
bloqueio	integração	tarifa
BR	interditado	táxi
buraco	lancha	trânsito
buracos	moto	travessia
calçada	motocicleta	travessias
calçadas	multa	vale-transporte
caminhão	ônibus	veículo
carro	paquímetro	veículos
ciclovia	passarela	velocidade
congestionamento	passarelas	via
deslocamento	rodovia	
engarrafamento	rodoviária	

Tabela C.9: Palavras Sementes para Saúde.

agente	clínicos	hepatite	oncologia
aids	colesterol	hiv	papiloma
alcool	combate	hospitais	peste
Alcoolismo	comunitário	hospital	posto
Anemia	dengue	hospitalar	preservativo
Anorexia	diagnostico	hpv	prevencao
Anti-retroviral	digestao	imunológica	raiva
Antibiótico	doença	imunológico	sanitaria
asma	doenças	influenza	saúde
bronquite	Embriaguez	leito	sexualidade
camisinha	endemia	Leptospirose	sifilis
campanha	endemias	lesão	tabagismo
câncer	epidemia	malaria	terapeutica
cancro	epidemias	medicamento	tétano
candidiase	exame	medicamentos	tuberculose
cardiovasculares	exames	médico	vacina
catarata	febre	médicos	vacinação
cigarro	gripe	mental	vacinas
Citopatologia	h1n1	nutricao	
clínico	hemocentro	obesidade	

Tabela C.10: Palavras Sementes para Tecnologia & Inovação.

algoritmo	criação	INTELECTUAL	projetos
análise	criptografia	invencao	prototipo
APL	criptografico	metodologia	qualificação
aplicada	dados	metodos	resultados
arranjo	desenvolvimento	modelo	SEBRAE
automação	digital	novidade	sistema
base	engenharia	OCEANTEC	sistemas
bioinformatica	engenho	parque	START-UP
biometria	estatistica	patente	STARTUP
BIOTECNOLOGIA	gestao	pesquisa	tecnologia
Certificação	hardware	planejamento	tecnologias
ciencia	incubacao	PÓS-INCUBAÇÃO	tecnológico
cientifico	incubadora	processamento	treinamento
cluster	industria	processo	WORKSHOP
clusters	industrial	processos	
consultoria	inovacao	profissional	
cps	inovador	projeto	

Tabela C.11: Palavras Sementes para Urbanismo.

acessibilidade	canalete	ecodesenvolvimento	onibus	saneamento
acesso	canaletes	ecossistema	organica	sanitario
acostamento	canteiro	edificação	parque	Saraiva
ALTIMETRIA	cartografia	edifício	parques	seletiva
ambiental	cassino	enchentes	passarela	Tamandaré
arquitetura	ciclista	Ferreira	passarelas	terreno
arroio	ciclo	habitacao	patrimonio	topografia
arraumento	ciclofaixa	ibge	pavimentacao	topologia
árvore	ciclovia	impacto	pavimento	travessia
assentamento	cidade	infraestrutura	pedestre	travessias
assentamentos	circulacao	jardim	pedestres	turismo
aterro	coleta	lagoa	perimetro	urbana
bairro	conservacao	largo	população	urbanismo
barra	corredor	limpa	praça	urbanistica
barreiras	demolição	limpeza	praças	urbanisticas
básico	densidade	lixo	praia	urbano
beco	desapropriacao	loteamento	quartel	Xavier
biciletario	descentralizacao	malha	reciclagem	zoneamento
bolaxa	desenvolvimento	mangueira	reforma	
bueiros	desterritorialização	metropolitana	revitalizacao	
cadeirante	drenagem	molhes	rodoviária	
calçamento	dunas	ocupacao	rural	

APÊNDICE D – EXPERIMENTOS REALIZADOS

Tabela D.1: Experimento I.

Aspecto	Experimento	DR¹	DNR²	VP³	FP⁴	VN⁵	FN⁶	Precisão	Revocação	F-score	Acurácia
Economia	representativas	25	217	15	7	210	10	6.818	6.000	6.383	9.298
Educação	representativas	49	193	40	7	186	9	8.511	8.163	8.333	9.339
Empreendedorismo	representativas	4	238	0	0	238	4	0.000	0.000	0.000	9.835
Energia	representativas	5	237	4	0	237	1	10.000	8.000	8.889	9.959
Governanca	representativas	23	219	18	16	203	5	5.294	7.826	6.316	9.132
Meio Ambiente	representativas	23	219	14	3	216	9	8.235	6.087	7.000	9.504
Mobilidade	representativas	17	225	4	2	223	13	6.667	2.353	3.478	9.380
Saúde	representativas	14	228	8	2	226	6	8.000	5.714	6.667	9.669
Segurança	representativas	67	175	60	9	166	7	8.696	8.955	8.824	9.339
Tecnologia & Inovação	representativas	4	238	0	0	238	4	0.000	0.000	0.000	9.835
Urbanismo	representativas	11	231	3	2	229	8	6.000	2.727	3.750	9.587
Economia	sementes	25	217	15	10	207	10	6.000	6.000	6.000	9.174
Educação	sementes	49	193	38	20	173	11	6.552	7.755	7.103	8.719
Empreendedorismo	sementes	4	238	0	1	237	4	0.000	0.000	0.000	9.793
Energia	sementes	5	237	3	0	237	2	10.000	6.000	7.500	9.917
Governanca	sementes	23	219	13	19	200	10	4.062	5.652	4.727	8.802
Meio Ambiente	sementes	23	219	9	0	219	14	10.000	3.913	5.625	9.421
Mobilidade	sementes	17	225	8	3	222	9	7.273	4.706	5.714	9.504
Saúde	sementes	14	228	6	2	226	8	7.500	4.286	5.455	9.587
Segurança	sementes	67	175	35	4	171	32	8.974	5.224	6.604	8.512
Tecnologia & Inovação	sementes	4	238	1	2	236	3	3.333	2.500	2.857	9.793
Urbanismo	sementes	11	231	4	9	222	7	3.077	3.636	3.333	9.339
Economia	sementes e representativas	25	217	15	7	210	10	6.818	6.000	6.383	9.298
Educação	sementes e representativas	49	193	43	13	180	6	7.679	8.776	8.190	9.215
Empreendedorismo	sementes e representativas	4	238	0	0	238	4	0.000	0.000	0.000	9.835
Energia	sementes e representativas	5	237	4	0	237	1	10.000	8.000	8.889	9.959
Governanca	sementes e representativas	23	219	18	22	197	5	4.500	7.826	5.714	8.884
Meio Ambiente	sementes e representativas	23	219	14	1	218	9	9.333	6.087	7.368	9.587
Mobilidade	sementes e representativas	17	225	7	1	224	10	8.750	4.118	5.600	9.545
Saúde	sementes e representativas	14	228	8	0	228	6	10.000	5.714	7.273	9.752
Segurança	sementes e representativas	67	175	56	8	167	11	8.750	8.358	8.550	9.215
Tecnologia & Inovação	sementes e representativas	4	238	0	0	238	4	0.000	0.000	0.000	9.835
Urbanismo	sementes e representativas	11	231	4	1	230	7	8.000	3.636	5.000	9.669

¹ Documentos Relevantes, ² Documentos Não Relevantes,³ Verdadeiros Positivos, ⁴ Falsos Positivos,⁵ Verdadeiros Negativos, ⁶ Falsos Negativos

Tabela D.2: Experimento II.

Aspecto	Experimento	DR ¹	DNR ²	VP ³	FP ⁴	VN ⁵	FN ⁶	Precisão	Revocação	F-score	Acurácia
Economia	representativas	31	266	21	6	260	10	7.778	6.774	7.241	9.461
Educação	representativas	63	234	46	9	225	17	8.364	7.302	7.797	9.125
Empreendedorismo	representativas	3	294	0	0	294	3	0.000	0.000	0.000	9.899
Energia	representativas	6	291	4	0	291	2	10.000	6.667	8.000	9.933
Governanca	representativas	28	269	21	20	249	7	5.122	7.500	6.087	9.091
Meio Ambiente	representativas	28	269	23	7	262	5	7.667	8.214	7.931	9.596
Mobilidade	representativas	18	279	7	2	277	11	7.778	3.889	5.185	9.562
Saúde	representativas	16	281	11	1	280	5	9.167	6.875	7.857	9.798
Segurança	representativas	91	206	73	5	201	18	9.359	8.022	8.639	9.226
Tecnologia & Inovação	representativas	3	294	0	0	294	3	0.000	0.000	0.000	9.899
Urbanismo	representativas	10	287	3	1	286	7	7.500	3.000	4.286	9.731
Economia	sementes	31	266	16	12	254	15	5.714	5.161	5.424	9.091
Educação	sementes	63	234	46	21	213	17	6.866	7.302	7.077	8.721
Empreendedorismo	sementes	3	294	1	0	294	2	10.000	3.333	5.000	9.933
Energia	sementes	6	291	3	1	290	3	7.500	5.000	6.000	9.865
Governanca	sementes	28	269	15	25	244	13	3.750	5.357	4.412	8.721
Meio Ambiente	sementes	28	269	11	0	269	17	10.000	3.929	5.641	9.428
Mobilidade	sementes	18	279	6	5	274	12	5.455	3.333	4.138	9.428
Saúde	sementes	16	281	6	0	281	10	10.000	3.750	5.455	9.663
Segurança	sementes	91	206	51	3	203	40	9.444	5.604	7.034	8.552
Tecnologia & Inovação	sementes	3	294	0	0	294	3	0.000	0.000	0.000	9.899
Urbanismo	sementes	10	287	3	10	277	7	2.308	3.000	2.609	9.428
Economia	sementes e representativas	31	266	22	7	259	9	7.586	7.097	7.333	9.461
Educação	sementes e representativas	63	234	56	13	221	7	8.116	8.889	8.485	9.327
Empreendedorismo	sementes e representativas	3	294	0	0	294	3	0.000	0.000	0.000	9.899
Energia	sementes e representativas	6	291	3	0	291	3	10.000	5.000	6.667	9.899
Governanca	sementes e representativas	28	269	20	21	248	8	4.878	7.143	5.797	9.024
Meio Ambiente	sementes e representativas	28	269	21	2	267	7	9.130	7.500	8.235	9.697
Mobilidade	sementes e representativas	18	279	8	3	276	10	7.273	4.444	5.517	9.562
Saúde	sementes e representativas	16	281	11	0	281	5	10.000	6.875	8.148	9.832
Segurança	sementes e representativas	91	206	72	4	202	19	9.474	7.912	8.623	9.226
Tecnologia & Inovação	sementes e representativas	3	294	0	0	294	3	0.000	0.000	0.000	9.899
Urbanismo	sementes e representativas	10	287	4	1	286	6	8.000	4.000	5.333	9.764

¹ Documentos Relevantes, ² Documentos Não Relevantes,³ Verdadeiros Positivos, ⁴ Falsos Positivos,⁵ Verdadeiros Negativos, ⁶ Falsos Negativos

Tabela D.3: Experimento III.

Aspecto	Experimento	DR¹	DNR²	VP³	FP⁴	VN⁵	FN⁶	Precisão	Revocação	F-score	Acurácia
Economia	representativas	31	266	13	6	260	18	6.842	4.194	5.200	9.192
Educação	representativas	63	234	50	8	226	13	8.621	7.937	8.264	9.293
Empreendedorismo	representativas	3	294	0	1	293	3	0.000	0.000	0.000	9.865
Energia	representativas	6	291	3	0	291	3	10.000	5.000	6.667	9.899
Governanca	representativas	28	269	23	20	249	5	5.349	8.214	6.479	9.158
Meio Ambiente	representativas	28	269	24	7	262	4	7.742	8.571	8.136	9.630
Mobilidade	representativas	18	279	8	3	276	10	7.273	4.444	5.517	9.562
Saúde	representativas	16	281	8	3	278	8	7.273	5.000	5.926	9.630
Segurança	representativas	91	206	70	8	198	21	8.974	7.692	8.284	9.024
Tecnologia & Inovação	representativas	3	294	0	0	294	3	0.000	0.000	0.000	9.899
Urbanismo	representativas	10	287	5	1	286	5	8.333	5.000	6.250	9.798
Economia	sementes	31	266	16	14	252	15	5.333	5.161	5.246	9.024
Educação	sementes	63	234	42	16	218	21	7.241	6.667	6.942	8.754
Empreendedorismo	sementes	3	294	0	1	293	3	0.000	0.000	0.000	9.865
Energia	sementes	6	291	3	0	291	3	10.000	5.000	6.667	9.899
Governanca	sementes	28	269	19	34	235	9	3.585	6.786	4.691	8.552
Meio Ambiente	sementes	28	269	9	0	269	19	10.000	3.214	4.865	9.360
Mobilidade	sementes	18	279	6	7	272	12	4.615	3.333	3.871	9.360
Saúde	sementes	16	281	6	0	281	10	10.000	3.750	5.455	9.663
Segurança	sementes	91	206	43	2	204	48	9.556	4.725	6.324	8.316
Tecnologia & Inovação	sementes	3	294	2	1	293	1	6.667	6.667	6.667	9.933
Urbanismo	sementes	10	287	1	9	278	9	1.000	1.000	1.000	9.394
Economia	sementes e representativas	31	266	16	8	258	15	6.667	5.161	5.818	9.226
Educação	sementes e representativas	63	234	54	12	222	9	8.182	8.571	8.372	9.293
Empreendedorismo	sementes e representativas	3	294	2	0	294	1	10.000	6.667	8.000	9.966
Energia	sementes e representativas	6	291	3	0	291	3	10.000	5.000	6.667	9.899
Governanca	sementes e representativas	28	269	23	25	244	5	4.792	8.214	6.053	8.990
Meio Ambiente	sementes e representativas	28	269	24	1	268	4	9.600	8.571	9.057	9.832
Mobilidade	sementes e representativas	18	279	8	8	271	10	5.000	4.444	4.706	9.394
Saúde	sementes e representativas	16	281	7	1	280	9	8.750	4.375	5.833	9.663
Segurança	sementes e representativas	91	206	67	3	203	24	9.571	7.363	8.323	9.091
Tecnologia & Inovação	sementes e representativas	3	294	0	0	294	3	0.000	0.000	0.000	9.899
Urbanismo	sementes e representativas	10	287	5	3	284	5	6.250	5.000	5.556	9.731

¹ Documentos Relevantes, ² Documentos Não Relevantes,

³ Verdadeiros Positivos, ⁴ Falsos Positivos,

⁵ Verdadeiros Negativos, ⁶ Falsos Negativos

Tabela D.4: Experimento IV.

Aspecto	Experimento	DR ¹	DNR ²	VP ³	FP ⁴	VN ⁵	FN ⁶	Precisão	Revocação	F-score	Acurácia
Economia	representativas	31	266	19	5	261	12	7.917	6.129	6.909	9.428
Educação	representativas	63	234	46	7	227	17	8.679	7.302	7.931	9.192
Empreendedorismo	representativas	3	294	0	0	294	3	0.000	0.000	0.000	9.899
Energia	representativas	6	291	5	0	291	1	10.000	8.333	9.091	9.966
Governanca	representativas	28	269	21	23	246	7	4.773	7.500	5.833	8.990
Meio Ambiente	representativas	28	269	15	4	265	13	7.895	5.357	6.383	9.428
Mobilidade	representativas	18	279	8	3	276	10	7.273	4.444	5.517	9.562
Saúde	representativas	16	281	11	1	280	5	9.167	6.875	7.857	9.798
Segurança	representativas	91	206	76	15	191	15	8.352	8.352	8.352	8.990
Tecnologia & Inovação	representativas	3	294	0	0	294	3	0.000	0.000	0.000	9.899
Urbanismo	representativas	10	287	2	1	286	8	6.667	2.000	3.077	9.697
Economia	sementes	31	266	13	7	259	18	6.500	4.194	5.098	9.158
Educação	sementes	63	234	38	22	212	25	6.333	6.032	6.179	8.418
Empreendedorismo	sementes	3	294	0	1	293	3	0.000	0.000	0.000	9.865
Energia	sementes	6	291	1	1	290	5	5.000	1.667	2.500	9.798
Governanca	sementes	28	269	21	30	239	7	4.118	7.500	5.316	8.754
Meio Ambiente	sementes	28	269	7	0	269	21	10.000	2.500	4.000	9.293
Mobilidade	sementes	18	279	10	4	275	8	7.143	5.556	6.250	9.596
Saúde	sementes	16	281	8	1	280	8	8.889	5.000	6.400	9.697
Segurança	sementes	91	206	53	2	204	38	9.636	5.824	7.260	8.653
Tecnologia & Inovação	sementes	3	294	2	1	293	1	6.667	6.667	6.667	9.933
Urbanismo	sementes	10	287	3	13	274	7	1.875	3.000	2.308	9.327
Economia	sementes e representativas	31	266	22	3	263	9	8.800	7.097	7.857	9.596
Educação	sementes e representativas	63	234	49	11	223	14	8.167	7.778	7.967	9.158
Empreendedorismo	sementes e representativas	3	294	1	0	294	2	10.000	3.333	5.000	9.933
Energia	sementes e representativas	6	291	3	0	291	3	10.000	5.000	6.667	9.899
Governanca	sementes e representativas	28	269	22	27	242	6	4.490	7.857	5.714	8.889
Meio Ambiente	sementes e representativas	28	269	14	1	268	14	9.333	5.000	6.512	9.495
Mobilidade	sementes e representativas	18	279	8	2	277	10	8.000	4.444	5.714	9.596
Saúde	sementes e representativas	16	281	11	0	281	5	10.000	6.875	8.148	9.832
Segurança	sementes e representativas	91	206	74	6	200	17	9.250	8.132	8.655	9.226
Tecnologia & Inovação	sementes e representativas	3	294	0	0	294	3	0.000	0.000	0.000	9.899
Urbanismo	sementes e representativas	10	287	1	5	282	9	1.667	1.000	1.250	9.529

¹ Documentos Relevantes, ² Documentos Não Relevantes,³ Verdadeiros Positivos, ⁴ Falsos Positivos,⁵ Verdadeiros Negativos, ⁶ Falsos Negativos

Tabela D.5: Experimento V.

Aspecto	Experimento	DR ¹	DNR ²	VP ³	FP ⁴	VN ⁵	FN ⁶	Precisão	Revocação	F-score	Acurácia
Economia	representativas	31	266	15	1	265	16	9.375	4.839	6.383	9.428
Educação	representativas	63	234	50	8	226	13	8.621	7.937	8.264	9.293
Empreendedorismo	representativas	3	294	0	0	294	3	0.000	0.000	0.000	9.899
Energia	representativas	6	291	5	0	291	1	10.000	8.333	9.091	9.966
Governanca	representativas	28	269	24	17	252	4	5.854	8.571	6.957	9.293
Meio Ambiente	representativas	28	269	19	3	266	9	8.636	6.786	7.600	9.596
Mobilidade	representativas	18	279	7	1	278	11	8.750	3.889	5.385	9.596
Saúde	representativas	16	281	8	2	279	8	8.000	5.000	6.154	9.663
Segurança	representativas	91	206	80	20	186	11	8.000	8.791	8.377	8.956
Tecnologia & Inovação	representativas	3	294	0	0	294	3	0.000	0.000	0.000	9.899
Urbanismo	representativas	10	287	4	1	286	6	8.000	4.000	5.333	9.764
Economia	sementes	31	266	10	12	254	21	4.545	3.226	3.774	8.889
Educação	sementes	63	234	36	21	213	27	6.316	5.714	6.000	8.384
Empreendedorismo	sementes	3	294	0	0	294	3	0.000	0.000	0.000	9.899
Energia	sementes	6	291	2	0	291	4	10.000	3.333	5.000	9.865
Governanca	sementes	28	269	25	34	235	3	4.237	8.929	5.747	8.754
Meio Ambiente	sementes	28	269	10	0	269	18	10.000	3.571	5.263	9.394
Mobilidade	sementes	18	279	10	9	270	8	5.263	5.556	5.405	9.428
Saúde	sementes	16	281	1	0	281	15	10.000	0.625	1.176	9.495
Segurança	sementes	91	206	51	2	204	40	9.623	5.604	7.083	8.586
Tecnologia & Inovação	sementes	3	294	1	4	290	2	2.000	3.333	2.500	9.798
Urbanismo	sementes	10	287	3	10	277	7	2.308	3.000	2.609	9.428
Economia	sementes e representativas	31	266	18	5	261	13	7.826	5.806	6.667	9.394
Educação	sementes e representativas	63	234	50	14	220	13	7.812	7.937	7.874	9.091
Empreendedorismo	sementes e representativas	3	294	1	0	294	2	10.000	3.333	5.000	9.933
Energia	sementes e representativas	6	291	3	0	291	3	10.000	5.000	6.667	9.899
Governanca	sementes e representativas	28	269	26	19	250	2	5.778	9.286	7.123	9.293
Meio Ambiente	sementes e representativas	28	269	16	2	267	12	8.889	5.714	6.957	9.529
Mobilidade	sementes e representativas	18	279	7	5	274	11	5.833	3.889	4.667	9.461
Saúde	sementes e representativas	16	281	4	1	280	12	8.000	2.500	3.810	9.562
Segurança	sementes e representativas	91	206	73	11	195	18	8.690	8.022	8.343	9.024
Tecnologia & Inovação	sementes e representativas	3	294	0	0	294	3	0.000	0.000	0.000	9.899
Urbanismo	sementes e representativas	10	287	6	4	283	4	6.000	6.000	6.000	9.731

¹ Documentos Relevantes, ² Documentos Não Relevantes,

³ Verdadeiros Positivos, ⁴ Falsos Positivos,

⁵ Verdadeiros Negativos, ⁶ Falsos Negativos

Tabela D.6: Experimento VI.

Aspecto	Experimento	DR ¹	DNR ²	VP ³	FP ⁴	VN ⁵	FN ⁶	Precisão	Revocação	F-score	Acurácia
Economia	representativas	31	266	19	8	258	12	7.037	6.129	6.552	9.327
Educação	representativas	63	234	44	5	229	19	8.980	6.984	7.857	9.192
Empreendedorismo	representativas	3	294	0	0	294	3	0.000	0.000	0.000	9.899
Energia	representativas	6	291	4	0	291	2	10.000	6.667	8.000	9.933
Governanca	representativas	28	269	22	20	249	6	5.238	7.857	6.286	9.125
Meio Ambiente	representativas	28	269	21	2	267	7	9.130	7.500	8.235	9.697
Mobilidade	representativas	18	279	7	2	277	11	7.778	3.889	5.185	9.562
Saúde	representativas	16	281	7	1	280	9	8.750	4.375	5.833	9.663
Segurança	representativas	91	206	76	12	194	15	8.636	8.352	8.492	9.091
Tecnologia & Inovação	representativas	3	294	0	0	294	3	0.000	0.000	0.000	9.899
Urbanismo	representativas	10	287	5	1	286	5	8.333	5.000	6.250	9.798
Economia	sementes	31	266	19	12	254	12	6.129	6.129	6.129	9.192
Educação	sementes	63	234	38	15	219	25	7.170	6.032	6.552	8.653
Empreendedorismo	sementes	3	294	0	1	293	3	0.000	0.000	0.000	9.865
Energia	sementes	6	291	1	1	290	5	5.000	1.667	2.500	9.798
Governanca	sementes	28	269	18	29	240	10	3.830	6.429	4.800	8.687
Meio Ambiente	sementes	28	269	10	0	269	18	10.000	3.571	5.263	9.394
Mobilidade	sementes	18	279	11	6	273	7	6.471	6.111	6.286	9.562
Saúde	sementes	16	281	3	0	281	13	10.000	1.875	3.158	9.562
Segurança	sementes	91	206	46	2	204	45	9.583	5.055	6.619	8.418
Tecnologia & Inovação	sementes	3	294	1	2	292	2	3.333	3.333	3.333	9.865
Urbanismo	sementes	10	287	3	12	275	7	2.000	3.000	2.400	9.360
Economia	sementes e representativas	31	266	22	7	259	9	7.586	7.097	7.333	9.461
Educação	sementes e representativas	63	234	49	15	219	14	7.656	7.778	7.717	9.024
Empreendedorismo	sementes e representativas	3	294	0	0	294	3	0.000	0.000	0.000	9.899
Energia	sementes e representativas	6	291	1	0	291	5	10.000	1.667	2.857	9.832
Governanca	sementes e representativas	28	269	23	20	249	5	5.349	8.214	6.479	9.158
Meio Ambiente	sementes e representativas	28	269	17	2	267	11	8.947	6.071	7.234	9.562
Mobilidade	sementes e representativas	18	279	10	2	277	8	8.333	5.556	6.667	9.663
Saúde	sementes e representativas	16	281	5	0	281	11	10.000	3.125	4.762	9.630
Segurança	sementes e representativas	91	206	73	4	202	18	9.481	8.022	8.690	9.259
Tecnologia & Inovação	sementes e representativas	3	294	0	0	294	3	0.000	0.000	0.000	9.899
Urbanismo	sementes e representativas	10	287	3	5	282	7	3.750	3.000	3.333	9.596

¹ Documentos Relevantes, ² Documentos Não Relevantes,³ Verdadeiros Positivos, ⁴ Falsos Positivos,⁵ Verdadeiros Negativos, ⁶ Falsos Negativos

Tabela D.7: Experimento VII.

Aspecto	Experimento	DR¹	DNR²	VP³	FP⁴	VN⁵	FN⁶	Precisão	Revocação	F-score	Acurácia
Economia	representativas	31	266	19	3	263	12	8.636	6.129	7.170	9.495
Educação	representativas	63	234	44	4	230	19	9.167	6.984	7.928	9.226
Empreendedorismo	representativas	3	294	0	0	294	3	0.000	0.000	0.000	9.899
Energia	representativas	6	291	4	0	291	2	10.000	6.667	8.000	9.933
Governanca	representativas	28	269	25	18	251	3	5.814	8.929	7.042	9.293
Meio Ambiente	representativas	28	269	20	2	267	8	9.091	7.143	8.000	9.663
Mobilidade	representativas	18	279	9	2	277	9	8.182	5.000	6.207	9.630
Saúde	representativas	16	281	12	2	279	4	8.571	7.500	8.000	9.798
Segurança	representativas	91	206	75	12	194	16	8.621	8.242	8.427	9.057
Tecnologia & Inovação	representativas	3	294	0	0	294	3	0.000	0.000	0.000	9.899
Urbanismo	representativas	10	287	6	0	287	4	10.000	6.000	7.500	9.865
Economia	sementes	31	266	16	9	257	15	6.400	5.161	5.714	9.192
Educação	sementes	63	234	40	17	217	23	7.018	6.349	6.667	8.653
Empreendedorismo	sementes	3	294	0	1	293	3	0.000	0.000	0.000	9.865
Energia	sementes	6	291	3	1	290	3	7.500	5.000	6.000	9.865
Governanca	sementes	28	269	21	30	239	7	4.118	7.500	5.316	8.754
Meio Ambiente	sementes	28	269	8	0	269	20	10.000	2.857	4.444	9.327
Mobilidade	sementes	18	279	11	9	270	7	5.500	6.111	5.789	9.461
Saúde	sementes	16	281	8	1	280	8	8.889	5.000	6.400	9.697
Segurança	sementes	91	206	49	2	204	42	9.608	5.385	6.901	8.519
Tecnologia & Inovação	sementes	3	294	1	1	293	2	5.000	3.333	4.000	9.899
Urbanismo	sementes	10	287	3	12	275	7	2.000	3.000	2.400	9.360
Economia	sementes e representativas	31	266	21	8	258	10	7.241	6.774	7.000	9.394
Educação	sementes e representativas	63	234	50	8	226	13	8.621	7.937	8.264	9.293
Empreendedorismo	sementes e representativas	3	294	0	0	294	3	0.000	0.000	0.000	9.899
Energia	sementes e representativas	6	291	2	0	291	4	10.000	3.333	5.000	9.865
Governanca	sementes e representativas	28	269	26	21	248	2	5.532	9.286	6.933	9.226
Meio Ambiente	sementes e representativas	28	269	18	0	269	10	10.000	6.429	7.826	9.663
Mobilidade	sementes e representativas	18	279	12	7	272	6	6.316	6.667	6.486	9.562
Saúde	sementes e representativas	16	281	10	1	280	6	9.091	6.250	7.407	9.764
Segurança	sementes e representativas	91	206	72	5	201	19	9.351	7.912	8.571	9.192
Tecnologia & Inovação	sementes e representativas	3	294	0	0	294	3	0.000	0.000	0.000	9.899
Urbanismo	sementes e representativas	10	287	5	2	285	5	7.143	5.000	5.882	9.764

¹ Documentos Relevantes, ² Documentos Não Relevantes,

³ Verdadeiros Positivos, ⁴ Falsos Positivos,

⁵ Verdadeiros Negativos, ⁶ Falsos Negativos

Tabela D.8: Experimento VIII.

Aspecto	Experimento	DR ¹	DNR ²	VP ³	FP ⁴	VN ⁵	FN ⁶	Precisão	Revocação	F-score	Acurácia
Economia	representativas	31	266	20	4	262	11	8.333	6.452	7.273	9.495
Educação	representativas	63	234	53	5	229	10	9.138	8.413	8.760	9.495
Empreendedorismo	representativas	3	294	0	0	294	3	0.000	0.000	0.000	9.899
Energia	representativas	6	291	5	0	291	1	10.000	8.333	9.091	9.966
Governanca	representativas	28	269	23	11	258	5	6.765	8.214	7.419	9.461
Meio Ambiente	representativas	28	269	19	1	268	9	9.500	6.786	7.917	9.663
Mobilidade	representativas	18	279	6	4	275	12	6.000	3.333	4.286	9.461
Saúde	representativas	16	281	10	3	278	6	7.692	6.250	6.897	9.697
Segurança	representativas	91	206	74	8	198	17	9.024	8.132	8.555	9.158
Tecnologia & Inovação	representativas	3	294	0	0	294	3	0.000	0.000	0.000	9.899
Urbanismo	representativas	10	287	2	1	286	8	6.667	2.000	3.077	9.697
Economia	sementes	31	266	16	14	252	15	5.333	5.161	5.246	9.024
Educação	sementes	63	234	45	22	212	18	6.716	7.143	6.923	8.653
Empreendedorismo	sementes	3	294	0	1	293	3	0.000	0.000	0.000	9.865
Energia	sementes	6	291	1	0	291	5	10.000	1.667	2.857	9.832
Governanca	sementes	28	269	19	27	242	9	4.130	6.786	5.135	8.788
Meio Ambiente	sementes	28	269	11	0	269	17	10.000	3.929	5.641	9.428
Mobilidade	sementes	18	279	10	9	270	8	5.263	5.556	5.405	9.428
Saúde	sementes	16	281	8	1	280	8	8.889	5.000	6.400	9.697
Segurança	sementes	91	206	47	0	206	44	10.000	5.165	6.812	8.519
Tecnologia & Inovação	sementes	3	294	1	3	291	2	2.500	3.333	2.857	9.832
Urbanismo	sementes	10	287	1	18	269	9	0.526	1.000	0.690	9.091
Economia	sementes e representativas	31	266	22	4	262	9	8.462	7.097	7.719	9.562
Educação	sementes e representativas	63	234	55	13	221	8	8.088	8.730	8.397	9.293
Empreendedorismo	sementes e representativas	3	294	1	0	294	2	10.000	3.333	5.000	9.933
Energia	sementes e representativas	6	291	1	0	291	5	10.000	1.667	2.857	9.832
Governanca	sementes e representativas	28	269	26	20	249	2	5.652	9.286	7.027	9.259
Meio Ambiente	sementes e representativas	28	269	19	0	269	9	10.000	6.786	8.085	9.697
Mobilidade	sementes e representativas	18	279	10	9	270	8	5.263	5.556	5.405	9.428
Saúde	sementes e representativas	16	281	10	1	280	6	9.091	6.250	7.407	9.764
Segurança	sementes e representativas	91	206	70	5	201	21	9.333	7.692	8.434	9.125
Tecnologia & Inovação	sementes e representativas	3	294	0	0	294	3	0.000	0.000	0.000	9.899
Urbanismo	sementes e representativas	10	287	1	3	284	9	2.500	1.000	1.429	9.596

¹ Documentos Relevantes, ² Documentos Não Relevantes,³ Verdadeiros Positivos, ⁴ Falsos Positivos,⁵ Verdadeiros Negativos, ⁶ Falsos Negativos

Tabela D.9: Experimento IX.

Aspecto	Experimento	DR ¹	DNR ²	VP ³	FP ⁴	VN ⁵	FN ⁶	Precisão	Revocação	F-score	Acurácia
Economia	representativas	31	266	15	5	261	16	7.500	4.839	5.882	9.293
Educação	representativas	63	234	51	8	226	12	8.644	8.095	8.361	9.327
Empreendedorismo	representativas	3	294	0	0	294	3	0.000	0.000	0.000	9.899
Energia	representativas	6	291	3	0	291	3	10.000	5.000	6.667	9.899
Governanca	representativas	28	269	19	26	243	9	4.222	6.786	5.205	8.822
Meio Ambiente	representativas	28	269	21	1	268	7	9.545	7.500	8.400	9.731
Mobilidade	representativas	18	279	10	0	279	8	10.000	5.556	7.143	9.731
Saúde	representativas	16	281	8	4	277	8	6.667	5.000	5.714	9.596
Segurança	representativas	91	206	80	7	199	11	9.195	8.791	8.989	9.394
Tecnologia & Inovação	representativas	3	294	0	0	294	3	0.000	0.000	0.000	9.899
Urbanismo	representativas	10	287	4	2	285	6	6.667	4.000	5.000	9.731
Economia	sementes	31	266	10	14	252	21	4.167	3.226	3.636	8.822
Educação	sementes	63	234	39	19	215	24	6.724	6.190	6.446	8.552
Empreendedorismo	sementes	3	294	0	3	291	3	0.000	0.000	0.000	9.798
Energia	sementes	6	291	3	1	290	3	7.500	5.000	6.000	9.865
Governanca	sementes	28	269	19	34	235	9	3.585	6.786	4.691	8.552
Meio Ambiente	sementes	28	269	12	0	269	16	10.000	4.286	6.000	9.461
Mobilidade	sementes	18	279	8	3	276	10	7.273	4.444	5.517	9.562
Saúde	sementes	16	281	5	0	281	11	10.000	3.125	4.762	9.630
Segurança	sementes	91	206	51	3	203	40	9.444	5.604	7.034	8.552
Tecnologia & Inovação	sementes	3	294	2	3	291	1	4.000	6.667	5.000	9.865
Urbanismo	sementes	10	287	1	14	273	9	0.667	1.000	0.800	9.226
Economia	sementes e representativas	31	266	17	6	260	14	7.391	5.484	6.296	9.327
Educação	sementes e representativas	63	234	54	11	223	9	8.308	8.571	8.438	9.327
Empreendedorismo	sementes e representativas	3	294	0	0	294	3	0.000	0.000	0.000	9.899
Energia	sementes e representativas	6	291	2	1	290	4	6.667	3.333	4.444	9.832
Governanca	sementes e representativas	28	269	22	24	245	6	4.783	7.857	5.946	8.990
Meio Ambiente	sementes e representativas	28	269	20	1	268	8	9.524	7.143	8.163	9.697
Mobilidade	sementes e representativas	18	279	12	2	277	6	8.571	6.667	7.500	9.731
Saúde	sementes e representativas	16	281	8	3	278	8	7.273	5.000	5.926	9.630
Segurança	sementes e representativas	91	206	75	4	202	16	9.494	8.242	8.824	9.327
Tecnologia & Inovação	sementes e representativas	3	294	0	0	294	3	0.000	0.000	0.000	9.899
Urbanismo	sementes e representativas	10	287	2	2	285	8	5.000	2.000	2.857	9.663

¹ Documentos Relevantes, ² Documentos Não Relevantes,³ Verdadeiros Positivos, ⁴ Falsos Positivos,⁵ Verdadeiros Negativos, ⁶ Falsos Negativos

Tabela D.10: Experimento X.

Aspecto	Experimento	DR ¹	DNR ²	VP ³	FP ⁴	VN ⁵	FN ⁶	Precisão	Revocação	F-score	Acurácia
Economia	representativas	31	266	20	3	263	11	8.696	6.452	7.407	9.529
Educação	representativas	63	234	45	7	227	18	8.654	7.143	7.826	9.158
Empreendedorismo	representativas	3	294	0	0	294	3	0.000	0.000	0.000	9.899
Energia	representativas	6	291	4	0	291	2	10.000	6.667	8.000	9.933
Governanca	representativas	28	269	25	19	250	3	5.682	8.929	6.944	9.259
Meio Ambiente	representativas	28	269	20	5	264	8	8.000	7.143	7.547	9.562
Mobilidade	representativas	18	279	8	3	276	10	7.273	4.444	5.517	9.562
Saúde	representativas	16	281	10	2	279	6	8.333	6.250	7.143	9.731
Segurança	representativas	91	206	77	9	197	14	8.953	8.462	8.701	9.226
Tecnologia & Inovação	representativas	3	294	0	0	294	3	0.000	0.000	0.000	9.899
Urbanismo	representativas	10	287	2	0	287	8	10.000	2.000	3.333	9.731
Economia	sementes	31	266	16	11	255	15	5.926	5.161	5.517	9.125
Educação	sementes	63	234	40	22	212	23	6.452	6.349	6.400	8.485
Empreendedorismo	sementes	3	294	1	3	291	2	2.500	3.333	2.857	9.832
Energia	sementes	6	291	3	1	290	3	7.500	5.000	6.000	9.865
Governanca	sementes	28	269	18	26	243	10	4.091	6.429	5.000	8.788
Meio Ambiente	sementes	28	269	10	0	269	18	10.000	3.571	5.263	9.394
Mobilidade	sementes	18	279	6	5	274	12	5.455	3.333	4.138	9.428
Saúde	sementes	16	281	5	1	280	11	8.333	3.125	4.545	9.596
Segurança	sementes	91	206	49	3	203	42	9.423	5.385	6.853	8.485
Tecnologia & Inovação	sementes	3	294	0	3	291	3	0.000	0.000	0.000	9.798
Urbanismo	sementes	10	287	1	6	281	9	1.429	1.000	1.176	9.495
Economia	sementes e representativas	31	266	21	6	260	10	7.778	6.774	7.241	9.461
Educação	sementes e representativas	63	234	50	9	225	13	8.475	7.937	8.197	9.259
Empreendedorismo	sementes e representativas	3	294	0	0	294	3	0.000	0.000	0.000	9.899
Energia	sementes e representativas	6	291	3	1	290	3	7.500	5.000	6.000	9.865
Governanca	sementes e representativas	28	269	28	23	246	0	5.490	10.000	7.089	9.226
Meio Ambiente	sementes e representativas	28	269	18	1	268	10	9.474	6.429	7.660	9.630
Mobilidade	sementes e representativas	18	279	10	3	276	8	7.692	5.556	6.452	9.630
Saúde	sementes e representativas	16	281	9	0	281	7	10.000	5.625	7.200	9.764
Segurança	sementes e representativas	91	206	73	6	200	18	9.241	8.022	8.588	9.192
Tecnologia & Inovação	sementes e representativas	3	294	0	0	294	3	0.000	0.000	0.000	9.899
Urbanismo	sementes e representativas	10	287	2	0	287	8	10.000	2.000	3.333	9.731

¹ Documentos Relevantes, ² Documentos Não Relevantes,³ Verdadeiros Positivos, ⁴ Falsos Positivos,⁵ Verdadeiros Negativos, ⁶ Falsos Negativos

APÊNDICE E – TERMOS REPRESENTATIVOS A PARTIR DO DATASET ANOTADO

Tabela E.1: Termos Representativos para o aspecto Economia. Parte 1.

#BAURU	artesanais	capacidade	Conciliação	degustar	empresários
#Bauruzinho	artesanato	capacitação	concretos	Delivery	empresas
#ComidaCaseira	assistente	Capitania	condeiteiro	demissão	empréstimos
#Delicias	atendimento	carbono	confia	Demissões	encanador
#DeliciososLanches	atualizações	carbono/nitrog	Confiança	DEMITE	encomenda
#DogÉShow	atum	Cardápio	conformidades	densidade	encomendar
#LancheMaravilhoso	aumenta	Carne	conscientização	desconto	encomende
#PartiuDog	automação	carpinteiro	Conselho	Desemprego	ensino-aprendizagem
#VemProDog	autônomo	carreira	consequência	DESMONTAGEM	entradas
abono	balanço	cartão	Consertos	desperdícios	entrega
achocolatados	balconista	cartões	construção	destinos	entregas
acompanhamentos	bancários	cascos	construções	Dezenas	equipamentos
açougueiro	banco	categoria	Construtora	DIÁLOGOS	Ervilha
adesão	banco	cdl	CONSULTORA	DIARIAMENTE	escavadeira
adiantamento	bancos	cenoura	consultoras	DIFERENCIAS	especialidade
administradores	banners	centrífugas	consumidores	dinheiro	especiarias
agência	bares	cerveja	consumo	DINHEIRO	espécie
AGÊNCIAS	barris	cervejas	container	DISPONÍVEIS	Esprírito
agrotóxicos	base	cestas	contas	Dívida	estabelecimento
água	batata	chamada	conteúdo	dívida	Estação
ajudante	bauru	chapista	contextos	dívidas	estaleiro
Alaminuta	baurú	chefia	contextualização	Docas	estaleiros
alface	baurus	chimarrão	contratação	dólares	estigma
alho	BAURUZINHO	Chocolate	CONTRATE	Economia	estoque
alimentação	bebida	Churrasco	contrato	ecovix	estranhos
alimentação	benefícios	classe	controla	editorial	estratégia
alimento	Bife	Cliente	convés	efeitos	estudos
alimentos	Bioflocos	coach	corte	eficiência	estufas
Almoce	bolinho	COACHING	COSMÉTICOS	efluentes	EUCALIPTO
almoço	bolsa	Colaboração	costura	eletricista	excretos
altera	bolso	colaboradores	costuras	eletrodomésticos	Excursão
alto-mar	bombas	colarinho	costureira	eletronica	Exemplos
aluguel	bordo	Comercial	couro	Eletrotudo	expectativa
alumínio	Boutique	comercialização	cozinha	ELITE	expectativas
Ambiente	brócolis	comerciante	CPRS	EMA/FURG	exponham
amônia	buffet	comerciante	credibilidade	EMATER	exportação
andaimes	burocracia	comerciário	crédito	embarcações	extras
antecipadamente	caçambeiro	Comerciários	credores	emissão	EXTRATO
APL	cachorro quente	comércio	crescimento	empreendedor	extratos
aplicação	cadeiras	comércios	CRISE	empreendimento	fábrica
apoiodor	CAIXA	comida	crueldade	empregada	falência
apressar	caixas	comida	CTPS	empregado	falha
APROVEITEM	calabresa	Comida	cuiá	empregador	falhas
Aquacultura	camarão	comidinha	culpa	empregador	farinha
aquaviários	CAMARÃO	comissário	cultivo	empregadores	farmácia
Ar-condicionado	caminhão	comitiva	currículo	empregados	fatura
Arena	caminhos	complementação	Currículos	emprego	feira
armador	camomila	compostos	custo	emprego	Feira
arranjo	campanheiros	Compra	débito	empregos	feirinha
arroz	campeiro	COMPRANDO	deferimento	empresa	Fenaban

Tabela E.2: Termos Representativos para o aspecto Economia. Parte 2.

Ferramentas	horizontes	logística	oportunidade	Porção	referência
ferreiro	Hotel	loja	OPORTUNIDADE	porções	referendo
fertilizante	idiomas	loja	organizacional	porteiro	Reformas
fertilizantes	Igressos	Loja	ouro	Porto	refrigeração
FGTS	illegalidade	lotéricas	ovos	portos	refrigerante
fidelidade	Ilha	lubrificador	P-58	Portos	regra
Fila	imitações	LUCRO	P-74	pós-paga	reinserção
foto	impacto	maçariqueiro	P-75	pós-pago	reivindica
folhas	Imposto	Macarrão	P-77	pós-venda	reivindicações
fome	indenização	Macarronada	pacote	pratos	Reivindicam
Força	Indivíduo	Manifestação	Padaria	PRÉ-FABRICADAS	relacionamento
força-tarefa	indumentárias	Maquinário	PAGA	Pré-Sal	relações
formação	Indústria	marcas	pagamento	Preço	relatório
formigas	industria	MARGENS	pagar	presidentes	renda
fornecedores	indústria	marmitas	paralisação	prestação	RENDA
fornecimento	indústrias	Marmitex	parceiro	princípios	reposição
fragmento	infinito	Master	parcela	Processos	repositor
fragmentos	Ingredientes	matérias	partes	produção	Rescisão
FRANQUEADO	Ingresso	MERCADO	participantes	produtividade	Reservas
franquia	iniciante	mercadorias	Patrão	produtivo	ressalto
freezers	Inovação	metalúrgico	patrões	produtores	ressocialização
FRITAS	inox	Micro-ondas	Peça	professor	restaurante
frutas	inteiros	milho	pendências	profissão	restaurantes
funcionamento	INTERNALIZAR	Milho	percepção	Profissional	resultados
funcionário	investidor	moído	perfumaria	Programa	retribuição
fundamento	investidores	Molhos	perito	promoção	retroescavadeira
fundos	Investimento	MONTADOR	PERSONALIZADO	promoções	REVENDEDORA
GANHOS	investimento	moradia	perspectiva	prontuários	revogatório
Garçons	iscas	motoniveladora	pesca	Pronúncia	RH
gastronomia	isenção	MÓVEL	pescador	proprietário	risco
gatos	itens	naval	pescadores	proteína	RODÍZIO
geladeira	janta	navios	pescados	Protesto	roteiro
Geleias	jantar	negociações	pesquisa	publicação	roubo
geração	Jornada	negócio	Pessimista	publicidade	Safra
GERAL	jugamento	negócio	petroleiro	purês	saídas
Gerenciais	Juros	níveis	Petroleiros	qualificação	salada
gerente	lanche	nível	PETRÓLEO	qualificação	salariais
gigante	lanche	norma	PIB	quantidade	salário
Governo	lancheria	NOSSO	pilchas	quebra	Salário
Governo	LANCHES	nota	pintura	quentinha	sapataria
Gramado	lanchonete	novidades	PIONEIRO	químico	saque
greve	laudo	nutricionista	pirataria	ração	SEBRAE
greve	Lava	oferecendo	piso	REAIS	secadora
grevistas	lazer	oferta	pizza	receita	segmento
guindastes	leilão	óleo	plataforma	receitas	Segurança
hamburguer	Leite	olho	plataformas	recuperação	Seguranças
hectare	liminar	opera	POBRE	recurso	Seguridade social
hierarquia	limitação	operador	POLÍTICOS	REDECARD	seguro
Hipercard	Local	Operadora	pólo	redução	Seguro
Hora-extra	locução	operadores	ponto	reestruturação	seguro-desemprego

Tabela E.3: Termos Representativos para o aspecto Economia. Parte 3.

senado	sindicatos	suflê	tensão	troca	vendedores
serralheiro	Sindilojas	supermercado	terminal	vaga	vender
serviço	SINE	supervisor	tolerância	vagas	veneno
serviços	SIRQ	sushi	TOMATE	Valor	ventilação
Servidor	sistema	tábua	tonelada	Valores	verbas
SESI	situação	tanque	toneladas	valorização	vigente
Shopping	SMDP	tanques	TRABALHADOR	varredor/capinador	viveiros
Shopping	sobremesa	taxa	trabalhadoras	vazamentos	voltam
show	sobrevivência	taxa	trabalhadores	veiculação	Whats
SindiBancários	soja	táxi	trabalhem	venda	WORKSHOP
sindicalistas	soldador	tecnologias	tradicionalistas	vendas	Xis
sindicância	solidariedade	tecon	transporte	vende	zooctenista
Sindicato	solidarizo	tele-entrega	treinamento	vendedor	Zoológico
sindicato	solo	Tele-Entrega	TREINO	vendedor	
Sindicato	sucos	teles	tristeza	vendedora	

Tabela E.4: Termos Representativos para o aspecto Educação. Parte 1.

#MundoMoinhoLiteratura	arquiteta	biblioteca	CiDEC-Sul	contrabaixo	diálogo
#MundoMoinhonaFearg	arquitetos	Bibliotecas	Cinquentenário	Conversa	dificuldade
#MundoMoinhoRecomenda	ARQUITETURA	Bicharada	cinturão	converte	DIGITAL
#RioGrande280anos	arte	Blog	Cirandarte	coordenação	direitos
#VitraisContosdoInvitro	arte-educadores	bola	circo	coordenador	diretivas
42FeiradoLivrodaFurg	artesanais	bolas	círculo	Coordenadoras	Diretor
AABB	Artesanatos	bolsista	circuitos	coordenadores	diretora
ABERTAS	ArtEstação	bolsita	classe	Coordenadoria	diretores
abordagem	artigos	boxe	Classes	COPA	diretoria
abraço	artilheiro	Brasileirão	clube	Coperse	Dirigentes
ACADÊMICA	artistas	brechó	co-autora	CORE	disciplinas
Acervo	ASKAB	brevemente	Co-orientador	coreografias	disputas
Acessibilidade	Assedio	brincadeiras	Co.place	Corpo	Dissertações
Ações	Assembleia	brindes	cobrança	correria	distribuição
acompanhamento	Assina	Cadete	colaboração	Cortinas	Diversidade
Adaptações	Astrofísica	CAIC	colação	Cosmologia	Divertida
adeptos	Ateliê	calendário	Coleção	costumes	divirta-se
adiamento	atividade	caminhadas	Colegial	Cotas	divisão
Administração	atividades	campeão	Colégio	Cpers	divulgar
Adolescente	Atletas	campeões	combate	CPIJ	docência
Adultos	atrações	campeonato	comédia	CRAS	doces
adversário	atrizes	campi	comediante	crença	domicílio
adversários	auditório	Campus	Comédias	crenças	doutoranda
Afro-Verão	Aula	Cancioneiros	comedouros	crescimento	Dreads
agenda	aulão	canto	competições	crítica	economista
agendamentos	Aulas	cantor	computador	cronograma	editais
álbum	Autor	Capes	computadores	curador	editora
alegria	autora	capitão	COMUNICADO	curiosidades	Educação
Alfabetização	autorais	caracterizações	concentração	curso	EDUCACIONAL
Alfabetizadores	autoral	carnavalescos	Concentrações	ursos	educador
alteração	autores	carreira	concerto	debate	Educandário
alterações	AUXÍLIO	carreiras	concessão	debates	Elaboração
alternativa	avalanche	Carreiros	conclusões	década	elasticidade
alternativo	Avaliação	carro-chefe	concurso	decreta-se	elemento
alunos	avaliadores	casamentos	concursos	déficits	EMEF
Ambiente	avanços	Cavalheiro	condições	delegação	EMEI
âmbito	Aventura	cenário	conduzir	democracia	emparelhou
âmbitos	bacharel	cênica	CONGRESSO	dependências	empate
analista	Bacharelado	certezas	conhecimento	Derrota	empatia
anfiteatro	bailarinos	chance	conhecimentos	desafios	encampado
anfitriões	baile	chances	CONSCIÊNCIA	descobertas	encerramento
antiguidade	banca	CIBER-FÍSICOS	CONSELHO	desconstrução	ênfases
Apatuf	bandas	ciclismo	conservação	descontentamento	engenharia
aperfeiçoamento	barcos	Ciclo	CONSTRUÇÃO	desempenho	Engenheiro
apoio	barroco	cidadã	contas	desfilar	Ensaio
apreciação	basquete	cidadania	conteúdo	desfile	ensinando
aprovação	Bateria	cidadão	conteúdos	Designer	Ensino
ARBITRAGEM	Beleza	CIDEC-Sul	contexto	desperdice	entretenimento
Arbitros	bênção	Ciência	contos	determinações	equipamento
arquibancadas	Berço	científicos	contra-ataque	dialogar	

Tabela E.5: Termos Representativos para o aspecto Educação. Parte 2.

equipes	fileiras	Igrejinha	Logistas	ofensas	pesquisador
escanteio	físico	Ilha	lotação	oferece	PIBITI/IFRS/CNPq
escolares	Fisioterapeutas	Ilhescas	Manicures	ofertando	pilates
escolinha	Flauta	IMEF	marciais	ofertar	Piquete
escritor	Flautas	imersão	marcial	Oficinas	pirâmide
escritora	fogos	IMPA	marítima	olhar	placar
escritoras	foliões	impressões	MASP	Operário	planejar
escritores	fomentar	inauguração	Matemática	opinião	Plano
especialista	força-tarefa	Incentivo	matrícula	oportunidade	plástica
especialistas	Formação	incidente	matrícululas	oportunidades	PNAIC
especializações	formato	incrições	Maturidade	Organizadora	poemas
esperanças	formatura	incursões	MCRG	organizadores	poesia
espetáculos	formaturas	Indígenas	mecânica	orgulho	poetas
ESPORTE	formulário	infraestrutura	medalhas	Orientação	politicagem
esportes	fotografias	INGLÊS	medalhista	orquestras	pontaria
esportivas	Fototeca	ingresso	meia-entrada	outorga	portais
Estação	Fronteira	Ingressos	memórias	Padroeira	Português
Estádio	fundação	Iniciação	mercados	página	Pós-graduação
estande	fundos	iniciantes	Mestrado	páginas	posições
Estatística	FURG	inscrever	mídias	pagode	povos
estilos	futsal	instituição	milagre	Palavras	Pré-Cálculo
estreante	ganhos	Instituições	miniauditório	palco	prédios
estréia	Garotos	instrutor	minicurso	palestra	premiação
ESTUDA	gauchão	intercâmbio	minicursos	palestras	prêmio
estudante	gaúchos	interessarem	ministrante	parabeniza	prêmios
estudantes	gênero	intervenções	modalidades	paradigmas	preparador
Estudantil	gêneros	Intolerância	monitor	paralisação	Preparativos
estude	Geografia	JECIRG	monitores	Paróquia	prestigar
estudo	gerenciador	joga	Mostra	participação	pretensão
evolução	GGI-M	jogador	mostras	participações	Primavera
examinadora	goleiro	Junina	motivação	participante	Procura
excelência	gols	jurados	Multipalco	participantes	produtora
excesso	gostos	kits	Municipais	Particulares	PROEN
experiência	graduação	lanterna	MUNICÍPIO	partida	professora
experiências	graffiti	largada	municípios	pastéis	professoras
experimentos	grafiteiro	lateral	Museus	Patrimônio	programar
exposição	grafiteiros	leitura	música	patrocinadores	projeção
Exposições	Gramado	leituras	músico	patrocínio	Proteção
expressão	Grande/CAIC	Lemes	músicos	patrono	protestos
expressões	Grande/IFRS	letivo	nacionalidades	pele	prova
Faculdade	GRATUITO	Libras	natalinos	pênalti	Provas
Fapergs	grau	Licenciatura	necessidades	Percurso	psicopedagoga
ferramenta	Grêmio	liderança	negociação	percussão	pública
Férrea	gremista	lideranças	Nepe	perfeição	qualidade
festeiros	guitarra	Liga	NEPE/FURG	periféricos	qualificação
festividade	hábito	Língua	NEPGS	periódico	Quintanilha
festividades	Horizonte	Literatura	Nível	Período	Quitanda
FGTRI	hospedagem	Livraria	nocaute	permanece	reaproveitamento
FIATECI	humor	livreiros	Núcleo	personagens	recepções
fibra	IFRS	livros	Núcleos	pescadoras	Recesso
filas	IFRS/Campus	locomoção	OBMEP	pesquisa	Recuperação

Tabela E.6: Termos Representativos para o aspecto Educação. Parte 3.

reflexão	Saudações	situação	Título	UFPel
reflexões	saudosa	SJN-RG	títulos	UNICAMP
regiões	SECRETARIA	SMCAS	TOCHA	universidades
Regulamento	secretário	SMEd	tomada	USP-SC
reivindicações	séculos	SMGA	Torcedor	USP/POLI
relacionamento	sede	SMTEL	torcer	utilize
REMAT	segmentos	SMTP	torcida	vagas
repreende	seleção	sobrancelhas	Toxicologia	validade
representante	Semana	sócio-torcedor	tradição	variedade
Representantes	semestrais	solemnidade	tradicionalista	velejadores
representativa	semestre	solicitações	tradutora	vencedores
respeita	semifinais	soluções	TRAJETO	Verão
respeito	Seminário	sonhos	trajetória	verdades
reuniões	seminários	sorteio	tranças	Veteranos
Revista	SENAT	submissão	transferência	Viajante
RGsurf	SENEGALESES	Substitutivas	transitam	vigentes
Rio-grandinos	sentimento	sucessos	transversalidades	vigor
ritmo	sentindo	SUSPENSAS	trapiche	vínculo
ritmos	Sentinela	talento	tratamentos	vínculos
ritual	sertanejo	talentos	trave	violão
rock	Servidores	teatrais	treinador	violoncelo
rodada	Sesi	teatro	Tributo	virtude
romance	SEST	TÉCNICA	tricampeão	visibilidade
Roteiro	SEST/SENAT	Técnicas	tricolor	visitação
rubro-verde	Sexo	Técnico	Trilha	visitações
Sacramentos	sexualidade	Tecnologia	troféus	visitante
sala	shows	tenores	turistas	vocal
salário	Sict	Terno	turísticas	volante/zagueiro
Salas	simpatia	texto	turma	Ypiranga
samba	Simpósios	textos	turmas	zagueiro
SAMPA	SINSC	time	turno	

Tabela E.7: Termos Representativos para o aspecto Empreendedorismo.

Aceleradora	Co.place	energéticas	franquias	Mercado	projetos
administração	comercializar	engenheiro	FURG	Métrica	promete
Advocacia	comerciantes	eólica	futuro	Métricas	qualidade
Afiliados	compliance	Escalabilidade	Gestão	Metro	recursos
Agenciamentos	Conferência	escritório	giro	mínimas	Remarketing
alimentos	consumidores	esforços	guia	negociações	Renda
alvará	contrato	estandes	importância	negócio	representante
alvarás	corporativo	estrutura	incubadora	negócios	resultados
aplicações	coworker	expectativa	indústria	Negócios	sanitário
apoio	Coworking	exposição	indústrias	Networking	SEBRAE/RS
Aporte	crise	exposições	Inovação	operações	Shopping
ARBRA	curso	expositor	inovadores	orçamento	situação
Arranjo	debates	Expositores	instituição	órgãos	soluções
atividades	desafio	Fazenda	instituições	Outsourcing	Stakeholders
B2B	desafios	fearg	Intraempreendedorismo	palestrantes	Startup
B2C	descumprimento	fecis	Investidor-anjo	palestras	Startups
bancas	dificuldades	feira	investimentos	parceria	Supply
bilionários	Diretor	feira	licença	Patente	tainha
Bootstrapping	diretora	feirantes	LMS	patrocinadores	Tecnologia
Break-even	Dirigentes	feiras	Lojistas	peixes	Tecnologias
Budget	Empreende	fejunca	lonas	perspectivas	turístico
Business	empreendedorismo	Feniartes	lucro	Pesca	venda
Capital	Empreendedorismo	festa	lucros	pescado	
Celulose	empreendimento	filés	manipulação	pescados	
cenário	empresa	filial	mar	planejamento	
CIDEC-SUL	empresarial	Follow-up	Marítimos	potencial	
Co-investimento	empresas	franquia	Marketing	produtores	

Tabela E.8: Termos Representativos para o aspecto Energia.

abastecimento	construção	eficiência	gasoduto	Líquidos	regaseificação
acessórios	construir	Efluentes	geração	matriz	secretário
ADEQUAÇÃO	consumidor	energética	gigawatts	meia	segredo
água	consumidoras	EQUIPAMENTOS	incentivos	Muita	serviços
ALERTA	Consumidores	esboço	indústrias	Nema	subestações
atração	consumo	estabilidade	iniciem	oportunizados	tópico
Bolognesi	contratada	ESTRUTURAS	Inovação	perspectivas	transmissão
caldeiras	contrato	fábrica	inspeções	pico	Triunfo
capacidade	contratos	fatura	instabilidade	píer	turbina
Captação	danos	Fepam	instalação	píeres	turbinas
CCEE	desabastecidos	Fiscalizações	instalações	POSTE	Usina
CEEE	desligamento	Fonte	irregularidade	potência	vapor
ciclo	desligamentos	fraudes	licença	presenças	
Companhia	distribuição	funcionamento	licenças	RECONDUTORAMENTO	
concessão	dólar	Furg	licenciamento	REDE	

Tabela E.9: Termos Representativos para o aspecto Governança. Parte 1.

#fora_os_dois	anúncio	cadastramento	combater	convenções	diminuições
ABALADAS	anúncios	cadastro	COMDEC	convênio	dinâmica
ABALADO	APAE	cadastrados	comercialização	convênios	dique
ABCX	Apátridas	cadelas	Comirat-RG	Convívio	diretores
abstenção	apelo	cafeteria	comissão	Coordenação	diretriz
abstenções	aprovação	calamidade	comitê	coordenador	diretrizes
acesse	apuração	calcário	comodidade	coordenadora	discussão
Acessibilidade	AQUICULTURA	calendário	Comparecimento	Coordenadoria	discussões
acesso	aquisição	CÂMARA	COMPDEC	CORPORATIVISTA	dívidas
acionamento	argumento	CAMBADA	COMPDEC/RG	cortada	doadora
ações	arrecadação	campanhas	competência	cortar	doadoras
acomodação	articulação	campos	complementação	creches	documento
acordos	Asdep	Canais	complexidade	crise	documentos
acusação	asilo	candidata	complexidades	CRLV	dólares
acusações	ASSEMBLEIA	candidatura	componentes	cronograma	doleiro
adequações	assembleias	candidaturas	composição	Cunha	ecocardiogramas
adesão	assinatura	canteiros	compras	CUSTARÃO	ecodopplers
adjunto	assinaturas	Capela	compromisso	custo	ecografias
administração	Assistência	caput	compromissos	custos	economista
administrações	associações	cargo	Comprovante	D'agostini	ECOSUL
adquirir	atacam	cartazes	Comunidade	dados	ECOVIX
Advocacia	atendimento	CARTEL	concessão	DATC	edital
advogada	atitude	Cartório	conclusão	debates	efeitos
ADVOGADOS	atos	cartórios	Concorrência	débitos	elaboração
AFASTAMENTO	atraso	casco	concorrentes	decisões	eleição
afirmações	atrasos	cassação	condicionado	declaração	eleitor
afirmar	atributos	castrações	configurações	Decreto	Eleitorado
Agasalho	atuação	CCMar	conluio	defasagem	eleitorais
agasalhos	Audiência	CDES	conquista	DEFINIÇÃO	eleitores
agenda	ausência	CELERIDADE	Conselho	delação	embasamento
Agradecimento	Autárquico	cemitério	Conselhos	deliberações	EMBRIAGUES
agravamento	automação	cenas	consequências	débitos	emergência
agressões	autoridades	chalé	constituição	decidam	empenho
agricultores	auxílio	chalés	construção	demandas	emplacamento
agricultura	avançar	chamamento	construções	democracia	empreendimentos
AHTS	bacias	chapa	construindo	DENUNCIA	empreiteiras
alinhar	bairro	chefe	construtoras	denúncias	empresariais
ALTERNATIVAS	bairros	chocolataria	consultor	DEPUTADO	empresarial
alunos	balneário	Cidadania	contábil	Deputados	empréstimo
Ambiente	bancário	cidadão	contabilidade	Desafios	encaminhando
ambientoterapia	barcaças	cidadãos	contadora	desajuste	endividamento
âmbito	bases	cidades	contas	desemprego	enfermagem
ambulatoriais	BASTA	Cinzas	contradições	desenvolvimento	engajamento
ameaças	benefício	cirurgia	contrapartida	desigualdade	entidade
amigos	bens	CNPJ	CONTRATAÇÕES	despesas	envelopes
ampliação	boicote	CNTE	contrato	detentos	envolvimento
andam	BOMBA	COAÇÃO	contratos	determinação	equipamentos
andamento	brasileiros	cobertores	contribuição	diferença	equipes
Anotações	Brigada	cofres	contribuintes	diferencia	Erradicação
antepenúltima	brigões	coibição	controle	dificuldades	escalão
anulação	BURACO	coligações	contrução	DIGNIDADE	escândalos

Tabela E.10: Termos Representativos para o aspecto Governança. Parte 2.

esclarecimentos	fortalecimento	inadimplência	LINDENMEYER	obra	podas
escolares	fórum	inauguração	linguagem	Obras	Poder
escolaridade	FPSO	incêndios	logística	obrigação	Política
escolas	frases	incentivo	lonas	obrigatoriedade	políticas
escritório	fraude	inclusão	lutar	OCEANTEC	políticos
Escritórios	fronteiras	inconformidade	luto	ofensiva	polo
escrevão	função	indicação	MAFIOSA	oncologia	Pombo
esfera	funcionalismo	índices	magnitude	OPoloNavalENosso	ponta
esforço	funcionamento	índole	mandado	oposição	Portal
egsoto	fundamentalismos	Infraestrutura	Mandato	ordens	pórtico
esquema	Fundiária	iniciativas	maquinás	orienta	prazos
estações	FURG	inquéritos	marcha	Ouvidoria	pré-cadastro
estaduais	gabinete	Inspetores	Marinheiros	P-71	prefeita
estadual	ganhadoras	instalação	MEC/FNDE	paciente	prefeito
estaleiro	ganhadores	instalador	medida	pagamento	prefeitos
estaleiros	garantem	intenções	medidas	palestrante	prefeituras
ESTATUTO	gastos	intensificação	megafones	palhaço	prejudica
estoques	geração	Interdição	meias-verdades	PAM-RG	prejuízo
estradas	Gestão	interpretação	melhorias	PAM/RG	prejuízos
estrangeiros	gestor	intervenção	menoridade	panfletagem	preocupação
Estratégia	gestores	INVESTIDORES	mentor	Parcelamento	PRESIDENTE
estupidez	GGI-M	investigações	Mercante	parcelamentos	prestação
ex-diretor	GGIM	Investigadores	mercantes	parceria	prestadores
ex-presidente	golpe	investimento	migrantes	PARCERIAS	prestígio
ex-presidentes	Golpistas	IPTU	mineiro	parlamentares	previdenciária
execução	Governador	IPVA	mineração	participação	prioritária
EXECUTIVO	governadores	JATO	minuta	partidos	prioritárias
executivos	governamentais	JUDICIÁRIO	Misericórdia	passagem	privatização
existência	governantes	juízes	mobilizações	passarem	procedimento
expectativa	Governo	julgarem	módulos	passistas	procedimentos
expediente	Governos	justificativa	monoboias	Patrimônio	PROCESSANTE
explicações	graminha	justificativas	monopólio	patrola	procurador-geral
exploração	greve	laboral	montantes	PCdoB	procuradores
exploradora	guincho	LADROES	morada	pedaços	profissionalizante
extras	guindaste	lágrimas	moradia	permissionária	promessa
fábricas	Habitação	Lava	moradias	Pesca	prometem
FÁCEIS	habitantes	Lava-Jato	MORO	Pescadores	prontuário
fachada	Hino	lavagem	MP-RS	Pesqueiro	pronunciamento
faculdades	hortifrutigranjeiros	legalização	multiplica	Petrobrás	propaganda
faixas	hoteleiro	legislação	Municipais	PETROLÃO	propriedade
Ferrovia	humildade	legislatura	MUNICIPAL	petroquímica	propinas
ferroviária	idealizadores	leis	Municipalidade	planilha	proponentes
FGTS	idiota	Libra	município	Plano	proposta
Finanças	imediações	Licença	NAÇÃO	PLANTÃO	PROSUS
fiscal	imobiliário	licenças	negociações	plantas	PROTECIONISMO
fiscalização	impeachment	Licenciamento	negócios	plataforma	protocolados
fiscalizações	impedimentos	licitação	Neoliberalismo	plataformas	protocolar
floricultura	imposto	licitações	nomeação	Platina	PROVIAS
Fome	impostos	licitado	nomes	pleito	providências
Fonte	improbidade	licitatório	NOMINAL	PMDB	prudencial
Forças	IMPUNIDADE	lideranças	obituários	PMRG	PSDB

Tabela E.11: Termos Representativos para o aspecto Governança. Parte 3.

PSDC	repartições	SAFADOS	SMHARF	Trabalhadores
PSOL	repassa	salarial	SMTEL	Trabalhista
PT-RS	REPASSAR	salários	soberania	transparência
química	repasse	Saneamento	solicitação	triagem
químicos	repasses	seções	solicitações	TRIBUNAIS
realidade	reposição	secretariado	solidariedade	tutores
realização	representação	Secretarias	soluções	urnas
rebocadores	representante	secretários	Sondas	Usina
recado	representantes	sede	sonho	valores
receita	repúdio	senador	sorveteria	veículos
recesso	requerimento	sensibilidade	subscritores	vencedores
recio	requerimentos	senso	subordinado	verbas
recondução	requisitos	Sérvia	Subprefeitura	vereadora
reconhecimento	rescisão	servidor	sucata	vereância
redemocratização	residente	servidoras	sucessória	vergonhas
reeleição	resolução	servidores	superintende	VESTUÁRIO
reestruturação	responsabilidade	sessões	superintendente	viajar
reflexo	ressonâncias	SETAL	Suprimentos	viaturas
reflexos	restrições	Setor	suspensão	vice
reformulação	resultado	setores	Talões	vice-prefeito
Refugiados	retrocesso	Shopping	tarifária	vice-prefeitos
regimento	reuniões	shoppings	TAXAS	vice-presidente
regras	réus	siderurgia	Tecnologia	Vices-presidente
regularização	revisão	sindical	telefônico	visitas
Relações	revolta	sistemas	telhas	votação
RELATOR	RIO-GRANDINO	situação	TEMER	votantes
remuneração	riquezas	situações	tentáculos	VOTO
rendimentos	Rodoviária	SMCAS	término	votos
Renova	ruas	SMCP	termos	xingam
renovação	RUMORES	SMDIER	TOLERÁ	YOUSSEF
renovações	rumos	SMDP	Torotama	ZOMBAM
reparo	safado	SMEd	TOYO	zonas

Tabela E.12: Termos Representativos para o aspecto Meio Ambiente. Parte 1.

#MataAtlântica	aves	concentração	emissão	instabilidade	nações
#recicleseushábitos	bairro	Conclusão	enchentes	intensidade	Nascente
#reduzaoconsumodeplásticos	baixas	condições	encostas	inundações	natureza
abate	barro	conscientização	entorno	isoladamente	navegação
abriga	beira	conservação	espécie	jacaré	nebulosidade
abrigue	beleza	consumo	espécies	kilometros	névoa
Abundância	belezas	continentes	Estação	lagartas	ninhos
ações	biodegradavel	Controle	estragos	lagartixa-das-dunas	normas
acúmulo	Biodiesel	Cooperação	estuário	Lagoa	nublado
advento	Biodiversidade	coordenador	estudo	Laguna	nutricionistas
afunda	biodiversidade	cordão	estudos	lama	nuvens
afundar	Biogás	corujas	exploração	latas	Oceanografia
agricultura	biomagnificaçao	costa	extinção	LCBC	oceanologia
Agrônomo	biomas	custou	Extrativista	licença	oceanos
agrotoxico	botos	danifica	fauna	Licenciamento	oeste
agrotoxicos	busquem	danos	Fazenda	LIMPA	Ondas
agrotóxicos	calor	danos	Fepam	liso	operação
água	camarão	decibelímetros	firme	lixo	operações
águas	camarão	declarações	Fiscalizações	lucro	organico
alagamentos	Carbonatos	decomposicao	flora	madeiramento	organicos
alcaloide	Carbono	Decompositores	Floresta	mamíferos	organismos
alcaloides	carências	defeso	Forestal	manejo	orla
alimentos	casca	degradação	fogo	mapas	painel
alugamento	cenários	denúncias	fontes	máquinas	paisagem
alturas	centros	departamentos	FOPEMEPE/RS	maranduvás	Pancadas
amanhecer	chove	depósito	fornecedor	maré	PAPELAO
ambiental	chover	desafio	fortalecimento	margens	parcerias
Ambientalistas	chuva	descargas	fósseis	marinha	parque
ambiente	chuva	deslizamentos	fragmentos	marinhos	Patos
Ambiente	chuvas	Desmatamento	frio	marítima	Patram
ambientes	chuvas	desmoronamentos	FURG	máxima	Patrimônio
analistas	chuvisqueiro	despejo	gama	MÉDIA	pautas
anfíbios	CICLONE	desperdício	gases	mediação	peixes
animais	cientistas	destine	geada	meteorológicas	perigo
anormalidades	cifra	detalhe	Genoves	Meteorologista	pesca
aparelhos	clima	diálogos	Governos	microesférulas	Pesca
aparições	Clima	Dióxido	granizo	micropartículas	pescado
APPs	climática	dragas	granizo	Microplásticos	pescadores
aproveitamento	climáticas	dunas	Habitat	milhas	pescados
Aquicultura	climático	Ecodesign	habitat	minima	pesquisa
Arborização	climáticos	Ecoeficiênciा	haste	Mitigação	pesquisadores
Área	Clorofluorcarbono	ecologia	hídricos	mm/dia	picos
areia	CO2	ecológico	Hotel	mm/h	plásticas
armazém	coberturas	ecológicos	Impacto	modificações	plastico
arrecada	coincide	ecossistema	impacto	Molhes	plásticos
árvores	coincidencia	ecossistemas	Impacts	Monumento	pluviais
atenuamento	Combustíveis	ecossistêmicos	impactos	mosquito	polêmica
atmosfera	COMPDEC/RG	Efluentes	incidência	mosquitos	polemizar
autorização	comportamento	embarcação	indicadores	mudas	Poluente
autos	composição	embargo	insetos	MUNICÍPIO	poluentes

Tabela E.13: Termos Representativos para o aspecto Meio Ambiente. Parte 2.

poluição	pulverizar	relatório	secretário	taturanas	UMIDADE
poluição	queimada	répteis	sedimento	temperatura	urbanização
populacionais	queimadas	réptil	sedimentos	TEMPERATURA	Vales
praias	químicos	Reserva	SENSAÇÃO	temperaturas	variação
precariedade	raios	resíduos	serração	temperaturas	vegetais
precaução	rajadas	resíduos	Situação	território	velocidades
prejuízos	rastilhão	resistência	situações	tesouro	ventania
preservacao	rastreamento	respeito	SMMA	torres	VENTO
preservação	ratos	responsabilidade	SNUC	Transgênico	Ventos
previsão	recuperação	ressaca	sol	transtornos	veraneio
Previsões	redes	Ressacão	solta	tratamento	viagens
primeira-dama	redução	rios	sonora	trator	vias
projeções	refeição	riscos	sudoeste	trovoada	virar
Proliferação	refeições	sacolas	supervisão	trovoadas	vizinhança
promete	Região	sacos	suspenção	tubinhos	
proporções	regiões	Safra	sustentabilidade	tuco-tuco	
Proteção	regulação	SEARG	Sustentável	turismo	

Tabela E.14: Termos Representativos para o aspecto Mobilidade.

ABERTA	caminhão	deslocamento	isenção	pista	SMMUA
acessibilidade	caminhões	destruição	lados	placa	SMTT
Acessibilidade	caminhoneiro	Detran-RS	lancha	placas	socorro
acesso	caminhoneiros	dividirá	linchar	PLANTÃO	subgrupo
accidentalidade	camionete	engarrafamento	lombada	policiais	subgrupos
acidente	caos	ERS	lombadas	Prefixo	subsídio
acidente	carga	espalharam	madrugada	Profissionalização	sufoco
acidentes	cargas	estacionamento	manobrar	proibir	tarifa
Acidentes	carona	estacionamento	manobras	Pronto-Socorro	tarifárias
ações	carreta	estada	MAPEIA	PRÓXIMO	tarifas
adequação	carretas	estrada	máquinas	publicitárias	táxi
adubo	carro	estudantes	margens	quilômetro	táxi
advogado	carro/carro	estudos	Masiero	quilômetros	terminais
ajuda	Carro/moto	etanol	migração	reconsideração	Trabalhador
ajude	CARROÇA	exceção	Mobilidade	reformulações	trabalhistas
ajustes	carroceiro	Fabril	moto	registrar	trafegam
ALERTA	CARROS	FERE	motocicleta	regulamentação	trafegar
AMBULÂNCIA	cavalo	ferimentos	motocicletas	Renovações	tráfego
armazenamento	cena	Fiscal	motociclistas	RESPEITA	tranqueira
aterro	centro/bairro	fiscalização	motofretistas	respeitarem	trânsito
atraso	choque	fluxo	Motor-Casa	retorno	trânsito
atravessar	ciclista	fluxo	Motorista	retrato	transportadoras
AUTO	ciclistas	fomentar	MOTORISTAS	retroportuários	Transportadores
automóvel	ciclovia	fraturas	motos	Riograndinos	transporte
autuação	cinto	frota	mototaxistas	rodovia	transportes
auxiliares	circulação	frotas	multa	rodoviária	Transposul
avenida	COLETIVO	Furg	NAVIO	Rodoviária	travessia
avenida	colisão	gasolina	óbito	rotatória	travessias
avenidas	colisões	gasolina	obras	rotatórias	trechos
avenidas	composto	gratuidade	ônibus	ROTULA	trem
BATE	compromete	gratuidades	onibus	RS-040	tributos
bicicleta	concentram	Graves	orientações	RS-118	TRIPULANTE
bloqueado	Condutor	guarda	página	RS-239	UPF-RS
bloqueia	condutores	guarda	paquímetro	ruas	usuários
bloqueio	congestionamento	guarnição	paradas	saibro	vale-transporte
BR	congestionamento	hipóteses	Paralisação	secretário	veículo
BR-116	contêineres	iluminação	param	semáforo	veículos
BR-392	controle	Imprudência	PARQUE	semáforo	velocidade
BR-471	correção	incremento	passageiro	semáforos	velocímetro
buraco	Cotista	indexador	passageiros	sensibilizaçao	verifica
buracos	críticos	indignação	passarela	sequelas	via
calçada	cruzamentos	infração	passarelas	sextas-feiras	viaduto
calçada	cuidado	infração	passe	sinal	viaja
calçadas	curva	instalação	patronais	sinalização	vias
calçadas	danificando	instantâneo	pedestres	Sinalização	vidas
cálculo	danos	integração	perímetro	Sindicam	vítima
caminhão	deslocamento	interditado	pico	situação	vítimas

Tabela E.15: Termos Representativos para o aspecto Saúde. Parte 1.

A/H1N1	caderneta	danos	glicemia	longevidade	portarias
A/H3N2	calamidade	demandas	graças	malaria	ponto
abandono	camisinha	demora	gripe	mama	preocupa
ações	campanha	dengue	gripe	Materno	preservativo
aditamento	Campanha	diagnosticos	h1n1	médica	preservativos
adolescentes	câncer	Diagnóstico	H1N1	medicação	pressão
Aedes	câncer	diagnósticos	hemocentro	medicações	prevencao
afrodescendentes	cancro	digestao	hepatite	medicamento	prevenção
agente	candidiase	dirigente	higiene	medicamentos	prevista
agentes	carcerário	distribuição	hiv	médico	prioritários
agravamentos	cardiologia	doação	HIV/AIDS	médico	procedência
aids	cardiologista	doações	hospitais	médicos	Pronto-socorro
AIDS	cardiovasculares	doadores	hospitais	medicos	próstata
alcool	carnes	doença	hospital	menores	proteção
Álcool	catarata	doenças	hospital	mental	pública
Alcoolismo	caxumba	dose	hospitalar	metodologias	público-alvo
alimentação	certificação	doses	hospitalar	micobactéria	puérperas
alunos	cigarro	doutor	Hospitalares	microbiologia	qualidade
Alzheimer	cirurgia	Drogas	hospitalizações	mortalidade	raiva
ambulâncias	cirurgias	Educação	hpv	mosquito	rede
ameaça	Citopatologia	eletrocardiograma	HU-Furg	Multivacinação	REFERÊNCIA
amostras	clinica	Embriaguez	HU-Furg/Ebserh	município	reforma
ampliação	clínicas	endemia	HU/Furg	Mycobacterium	RIOGRANDINA
Anemia	clínico	endemias	imunização	Neurologia	Rosa
Animal	clínicos	enfermagem	Imunizações	Neuros	sangue
Anorexia	coagulase	Enfermeiras	imunológica	nutricao	sanitaria
antecedencia	colesterol	Enfermeiros	imunológico	nutricionista	santa
Anti-retroviral	coliformes	epidemia	incidência	obesidade	Santas
Antibiótico	coloniais	epidemias	influenza	OBESIDADE	saúde
antibióticos	combate	equipamentos	Influenza	observa	SAÚDE
ANVISA	combate	Equipes	ingestão	oncologia	secretário
apoio	combater	especialistas	inspeção	orçamento	secretário-adjunto
aponta	complicações	esportes	integralidade	órgãos	segundo-secretário
apontam	compromiço	estabelecimentos	integridade	paciente	serviço
apresenta	comunitário	estética	internações	pacientes	servidores
asma	condições	estoque	internados	palestras	sifilis
assistência	Conselho	estoques	irresponsabilidade	palhaços	sífilis
ATENDIMENTO	contagens	estudos	IST/AIDS	papiloma	Sindisaúde
ATENDIMENTOS	contaminação	exame	IST/Aids/Hepatites	paredes	sintomas
atrasa	contrato	exames	ISTS	parto	situação
bactéria	Coordenadoria	expansão	laboratorial	penitenciárias	SMCAS
bairro	corpo	febre	laboratórios	perspectivas	subtipos
beneficiários	Cosems	físicos	Lacen	PESO	suplementos
biologia	CosemsRS	foco	leito	PESQUISA	tabagismo
Bipartite	cuidados	fragilidade	Leptospirose	peste	terapeutica
boca	cuidando	FURG	lesão	Plenária	terminar
breve	cumulo	gestantes	leve	pneumonias	tétano
bronquite	cura	getulio	liberdade	portadores	tipo

Tabela E.16: Termos Representativos para o aspecto Saúde. Parte 2.

titulares	UBSF	vacinal
transporte	UFPel	vacinas
Trânsito	unidade	vacinas
transmissão	unidades	veterinário
tratamento	urologista	Vigilância
tuberculose	vacina	vírus
TUBERCULOSE	vacina	visem
tuberculosis	vacinação	vistoria
tumor	vacinação	Vítima
tumores	vacinações	

Tabela E.17: Termos Representativos para o aspecto Segurança. Parte 1.

#ASSALTADO	antecedentes	bafômetro	CADEIA	cerveja	coordenador
1ªDP	aparelho	baiano	cães	chão	coronel
abate	aparência	balanças	CAÍDO	Charqueadas	coronhada
abdômem	apelido	balas	cair	chassi	coronhadas
abdômen	aperfeiçoamento	baleado	calça	chaves	corporação
abigeato	apetrechos	baleados	CALÇADÃO	chegará	corporal
abordá-lo	apocalipse	balneária	calcinha	choque	corpos
abordagens	apoio	balneários	calcinhas	chuveiro	corredor
abordar	apontar	BANDIDO	calçou	Chuvisca	corrente
abrigo	apreensão	bandidos	calibre	cicatrizes	corrupção
absinto	apreensões	banhista	calibres	cigarro	corsan
abuso	aproveitando-se	banhistas	calma	cigarros	cortada
ação	aquaplanou	barba	calmo	cinto	cortados
acertaram	areia	barranco	câmera	cintura	cortina
acerto	arma	barreira	caminhoneiro	circulação	costume
acertou	armas	barreiras	Caminhoneiros	circunstâncias	Cotegipe
acesso	arrastão	barriga	caminhonete	Civis	crack
acidente	ARREMESSAM	Batalhão	camiseta	cobertor	crime
acidentes	ARROMBADORES	bateu	Canastreiro	cobrador	crimes
acne	Arrombamento	batidas	canguru	cobrador	criminalidade
ações	arrombamentos	beber	canos	cocaína	criminosos
acontecimentos	arrombar	bebidas	capacete	cofre	cucuruto
acostamento	ASSALTA	beco	capinha	colaborador	cuidado
acusações	ASSALTAM	bermuda	capital	colchões	cunhada
Adolescente	assaltante	blitz	cápsulas	coletes	cunhado
adolescentes	ASSALTANTES	bloqueio	captura	colocam	custódia
adultos	assaltar	boletim	capuz	comandante	danificando
advertências	ASSALTARAM	bombeiros	característica	comandante-geral	danos
advogado	ASSALTO	bonde	carga	comarca	decomposição
afastamento	assaltos	botijão	Cargas	combate	decorrência
agência	assassinato	braços	carneadas	combater	DEFREC
agente	assassinatos	Briga	carnificina	combustível	DEGOLADA
Agentes	assassino	brigada	carregadores	comerciária	Delegacia
agredir	assassinos	brigadianas	carro/carro	companheiro	delegada
agressão	assusta	brigadiano	cartuchos	comparsa	delegado
agressões	ateando	brigadianos	casaco	comprimidos	delegados
agressor	Atestado	brincadeiras	casacos	comprometimento	delito
agressores	atingida	brutalidade	casarões	comunitária	delitos
agridam	atingido	bucha	catador	Comunitários	denúncia
alarme	atirando	buchholz	catraca	condutor	Denuncias
algemas	atribuição	buchinhas	catuaba	Condutores	denuncie
almirante	Atuando	Buchols	causadores	conduzir	Deodoro
alojamento	Audi/A3	bueiro	cavaleiros	confílio	deparam
aluga	automóvel	C.L.P	cela	conjunção	depoimento
alunos	autores	cabelos	celular	conserto	desaparecimento
alvo	autoridade	cabo	celulares	constata	descida
amadores	autoridades	cabra	centímetros	contingenciamento	desclassificação
amanhecer	autos	cachaça	Centro/Barra	contrabando	descoberta
ambulâncias	averiguação	cachorros	Certidão	contraprova	desconhecimento
Ameaça	avistar	cadeado	certidões	controle	desespero

Tabela E.18: Termos Representativos para o aspecto Segurança. Parte 2.

desobediência	estatura	frentistas	Incendiando	luzes	nacionalidade
destalhes	estilhaços	frete	indiciamento	macacão	namorada
detenção	estocada	frontal	indício	maconha	narcotráfico
detenções	estojo	fuga	indigestão	MAÇOS	necropsia
detento	estoura	FUGITIVO	indivíduos	Madeira	negros
dica	estudo	FUMO	inexistência	madrugada	Nenhum
digestão	ESTUPRADA	funcionalidade	inexistentes	malotes	Neves
discussão	ESTUPRADOR	Furg	infrator	mandado	ninja
dispararam	estupro	Furg/Carreiros	infratores	mandados	ninjas
disparo	ex-apenado	FURTADA	insegurança	mandante	notebook
Disparos	ex-companheira	furtadas	inspeções	mandantes	notebooks
disparou	execuções	Furto	inspetor	manta	numeração
distribuidores	executor	furto/abigeato	Inspetores	MARINHA	objeto
donos	exército	Furto/descuido	integração	martine	obstáculos
DPCA	exigir	furtos	integridade	máscaras	ocasiões
Dppa	êxito	fuzil	inteligência	massagem	ocorrência
DROGAS	explosão	garanta	internado	matagal	ocorrências
ecstasy	extase	garrafa	intuito	matar	óculos
efetividade	extintor	garrafadas	investigações	Mauá	ocupantes
efetua	extorsão	garrucha	Investigadores	maus-tratos	oferecem
efetuar	facada	garupa	irmão	mediana	operação
efetuarem	facadas	gaúchos	J.P.R.S	médias	operações
elementos	facão	gaveta	janela	medicação	orçamento
enfrentam	facas	GGI-M	janelas	meia-noite	ordem
entorpecentes	fachada	GIRL	jeans	meliantes	orientações
ENTREGADOR	fardamento	golpes	joalheria	MENOR	Padrão
ENTREGADORA	farmácias	GOLPISTAS	joelho	Menores	Paes
envenenamento	Faroeste	grade	joelhos	mensagens	palavrões
envólucros	FASE	grades	jogar	mercadoria	pânico
enxada	fatalidade	Grandino	jugamento	mercadorias	pano
equipes	feminicídio	gravar	JURI	minuano	papéis
ESCADA	ferimento	gravidade	JUSTIÇA	moça	papelotes
escapar	ferimentos	gritos	justiceiros	moças	parede
esconderijo	Ferrador	guanabara	ladra	mochila	partage
escoriação	ferragem	guanição	Ladrão	modo	passageira
escoriações	fezes	guarda	ladrões	moletom	passagens
esfaquear	fiança	guarita	lanchonetes	MOMENTO	passional
esforços	FILHO	guaritas	laranja	monitor	patrimônio
ESPANCANDO	fiquem	guarnição	laticínios	moradora	patrulhas
espingarda	fiscalização	guarnições	latinhas	morte	patrulheiros
espingardas	fita	Guimarães	latrocínio	mortes	pavor
esportistas	FLAGRADO	hidroviária	latrocínios	motivação	pedófilo
esposas	flanelinha	hipótese	lavagem	moto/carro	pedrada
espuma	foco	homicídios	lesão	Motoboy	pedras
espumante	foge	Hran	lesões	motoqueiro	pega
essse	folga	idades	licor	Mototaxista	peito
estabelecimento	foragidos	identificação	listras	MULHER	Pelotão
estabelecimentos	forças	iluminação	localização	munição	Penha
estaciona	formações	imprudência	logar	munições	penitenciária
estadia	frase	incêncio	louvor	muro	penitenciário

Tabela E.19: Termos Representativos para o aspecto Segurança. Parte 3.

perfurações	presenciar	receio	Robalo	suicídio	Tutelar
Perg	presidio	recepção	ROCAM	supermercado	uniforme
perícia	PRESOS	receptação	ROLETA	suspeitas	Urgência
periculosidade	pretensão	Receptador	rolo	suspeito	urina
perigo	prevenção	receptados	ronda	suspeitos	usuário
peritos	prisão	receptção	roubos	sustenta	vagabundos
pernas	prisional	reclusão	sacola	tactel	vagina
persegição	prisões	recompensa	salento	tatuagem	valas
personificação	procedência	reconhecimento	Salgado	táxis	valeta
pertences	produto	recusam	salvo-vidas	tecido	vandalismo
petecas	profilurb	redondezas	salvamento	tenente	Vândalos
pistola	profundidade	redução	salvamentos	tentativa	vantagens
pistolas	projéteis	reduto	Santuário	tequila	Vara
placa	promotor	REFÉM	Saveiros	terror	vazamento
PLANTÃO	Promotoria	referência	secretário	tesouraria	veiculo
Plaza	pronto-socorro	REFORÇO	segundos	testemunha	veículos
pneu	PRÓPRIA	refrigerantes	Segurança	testemunhas	veneno
Pneus	proprietária	refúgio	SEQUESTRA	tijolada	veranista
pochete	PRÓPRIO	regime	Sequestrador	tijolinhos	veranistas
polegadas	proteção	registram	sequestraram	tijolos	vestido
policia	pública	registros	SEQUESTRAVA	tiro	viatura
policial	PULAM	relatórios	sequestro	tiros	viaturas
polícias	punho	relógio	simulacro	tiroteio	Vice-Almirante
POPULARES	PÚRPURA	repercussão	sinal	tiroteios	VIDA
porta	quadrante	repetir	sinaleira	titular	videogame
porta-malas	quadril	represão	sinalizar	tomava	vidro
porta-voz	quadrilha	Reprodução	siri	tombador	vinculaço
Portaria	quadrilhas	repuxo	siri-azul	tórax	vingança
porte	Quartel	reservista	siris	touca	violação
posse	queijo	resgate	sistemas	toucas	violência
pracinha	queimaduras	responder	situação	Trabalhadora	vítima
pranchas	queixa	responsabilidade	situações	TRAFICANDO	vítimas
prata	quilômetro	resposta	skatistas	TRAFICANTE	vizinhos
prateado	QUILOS	ressarcimento	SMMUA	traficantes	vodka
praticar	rabeta	RESUMO	sobretudo	tráfico	vulnerabilidade
praxe	racha	retro	socorro	tragédia	whisky
prejudicar	racismo	revólver	socos	tranquilidade	
preocupam	radar	revólveres	substação	trânsito	
presa	Rapaz	rifle	substâncias	transporte	
presas	rapazes	riscos	suicidar	tumulto	

Tabela E.20: Termos Representativos para o aspecto Tecnologia & Inovação.

algoritmo	ciencia	digital	inovador	planejamento	SEBRAE
análise	cientifico	engenharia	INTELECTUAL	PÓS-INCUBAÇÃO	sistema
APL	cluster	engenho	invencao	processamento	sistemas
aplicada	clusters	estatistica	metodologia	processo	START-UP
arranjo	consultoria	gestao	metodos	processos	STARTUP
automação	cps	hardware	modelo	profissional	tecnologia
base	criação	incubacao	novidade	projeto	tecnologias
bioinformatica	criptografia	incubadora	OCEANTEC	projetos	tecnológico
biometria	criptografico	industria	parque	prototipo	treinamento
BIOTECNOLOGIA	dados	industrial	patente	qualificação	WORKSHOP
Certificação	desenvolvimento	inovacao	pesquisa	resultados	

Tabela E.21: Termos Representativos para o aspecto Urbanismo.

acabamentos	BALNEÁRIO	facilita	meio-fio	realinhamento	sustentação
ação	Barra	ferragens	meios-fios	realização	tela
Aceleração	brinquedo	floreiras	melhorias	recolocação	telas
Acessibilidade	brinquedos	Fundiária	monitores	reestruturação	terraplanagem
acesso	calçamento	FURG	mudas	reforma	terreno
acessos	Campanha	galpão	obra	Regularização	terreo
ações	ciclístico	gangorras	obras	remoção	transição
acomodação	colaboração	goleiras	organização	reparo	transtornos
administração	complexo	grama	pintura	reparos	trapiches
água	construção	Habitação	pisos	resíduos	travessa
alvenaria	construindo	habitacionais	pistas	retenção	tubulação
ambientes	contrata	harmonia	planejar	Rincão	tubulações
animais	cortaram	harmônica	Playground	Rodoviária	turnos
aparelhos	corte	harmonização	pluviais	roedores	uniformidade
apartamento	CRAS	higienização	poda	rótula	urbanização
apartamentos	cuida	iluminação	postes	ruas	vegetação
apropriação	danificação	implementação	prejudicam	Ruffier	vias
Arborização	depredação	Infraestrutura	programando	Saneamento	Viveiro
Arquipelágos	desembarque	inserção	projeta	Saraiva	wolff
árvores	Dragões	instala	projetos	secretário	zelo
assistentes	drenagem	instalação	propagação	Senandes	
autorização	educadores	instalações	PRORUAS	Serviços	
avenida	equipes	integração	proximidade	SMHRF	
avenidas	escoamento	lazer	psicólogos	superintendente	
balanços	estufas	Lixão	quadra	suportes	

APÊNDICE F – TERMOS REPRESENTATIVOS MAIS FREQUENTES A PARTIR DO DATASET ANOTADO



Figura F.1: Nuvem de palavras para o aspecto Economia.



Figura F.2: Nuvem de palavras para o aspecto Educação.



Figura F.3: Nuvem de palavras para o aspecto Empreendedorismo.



Figura F.4: Nuvem de palavras para o aspecto Energia.



Figura F.5: Nuvem de palavras para o aspecto Governança.



Figura F.6: Nuvem de palavras para o aspecto Meio Ambiente.



Figura F.7: Nuvem de palavras para o aspecto Mobilidade.



Figura F.8: Nuvem de palavras para o aspecto Saúde.

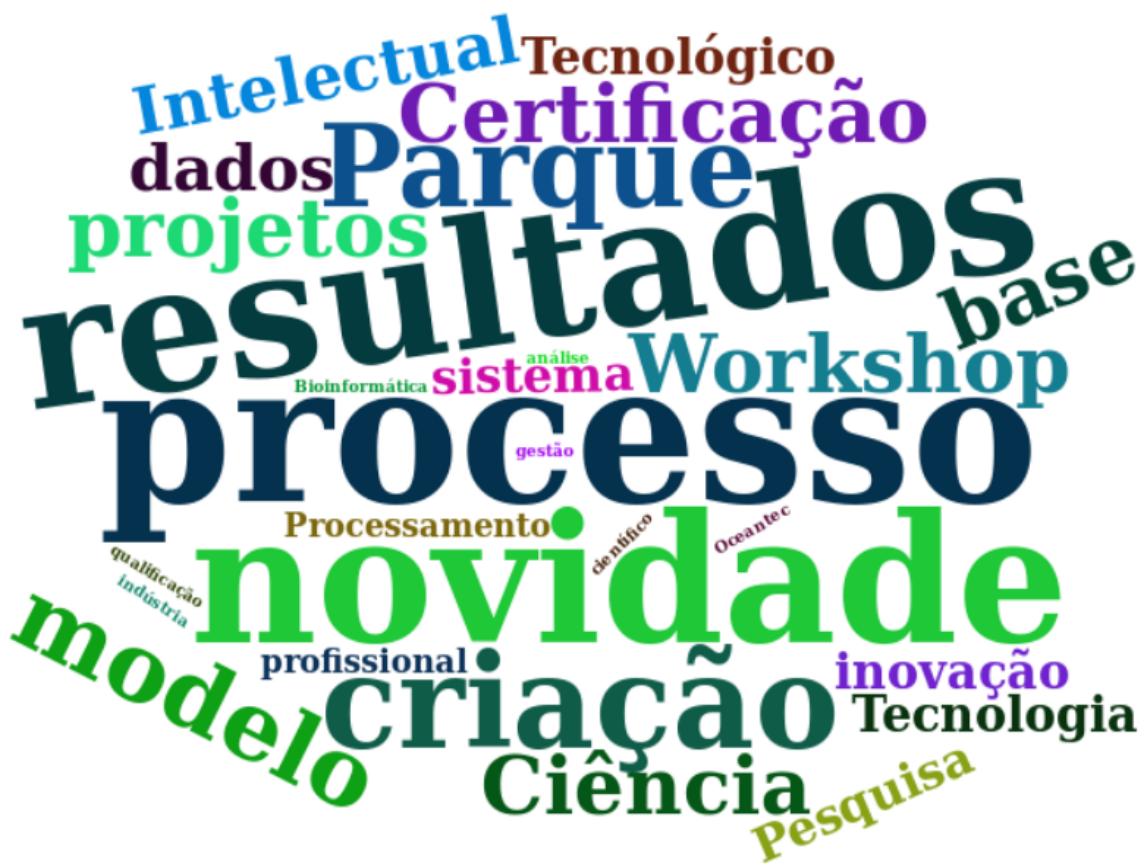


Figura F.9: Nuvem de palavras para o aspecto Tecnologia & Inovação.



Figura F.10: Nuvem de palavras para o aspecto Urbanismo.

APÊNDICE G – TERMOS REPRESENTATIVOS A PARTIR DE TODOS DOCUMENTOS CLASSIFICADOS

Tabela G.1: Termos Representativos para o aspecto Economia. Parte 1.

#BAURU	arranjo	capacidade	Conselho	DEMETE	entradas
#Bauruzinho	arroz	Capital	consequência	densidade	entrega
#ComidaCaseira	artesanais	Capitania	Consertos	desconto	entregas
#Delicias	assistente	carbono	construção	descontos	equipamentos
#DeliciososLanches	atendimento	carbono/nitrog	construções	DESMONTAGEM	Ervilha
#DogÉShow	ATRAÇÃO	Cardápio	Construtora	desperdícios	escavadeira
#LancheMaravilhoso	atualizações	carga	CONSULTORA	destinos	escolha
#PartiuDog	atum	Carnaval	consultoras	Dezenas	especialidade
#VemProDog	aumenta	Carne	Consumidor	DIÁLOGOS	especiarias
abono	aumentam	carpinteiro	consumidores	DIFERENCIAS	espécie
achocolatados	auxílio	carreira	consumo	dificuldade	Espírito
acompanhamentos	balança	cartão	contas	diminuição	estabelecimento
açougueiro	balanço	cartões	conteúdo	dinheiro	estabilidade
adesão	balconista	cascos	contextos	diretoria	Estação
adiantamento	Balneário	categoria	contextualização	discussão	Estado
adiantamentos	bancários	causas	CONTRATE	disposição	Estágios
administradores	bancos	cenoura	contrato	Dívida	estaleiro
agência	banners	centrífugas	contratos	dívidas	estaleiros
AGÊNCIAS	bares	cerveja	controla	dólares	estigma
agrotóxicos	barris	cervejas	controle	Economia	estoque
água	base	chamada	convés	editorial	estranhos
ajudante	batata	chapista	corte	efeitos	estratégia
Alaminuta	baurus	Chocolate	COSMÉTICOS	eficiência	estufas
alface	BAURUZINHO	Churrasco	costura	efluentes	etapas
alho	benefícios	classe	costuras	eletricista	EUCALIPTO
alimentação	Bife	Cliente	costureira	eletrodomésticos	excretos
alimento	bilionária	coach	couro	eletronica	Excursão
Almoce	bilionário	COACHING	cozinha	Eletrotudo	Executiva
almoço	Bioflocos	colaboradores	CPRS	ELITE	Exemplos
ALTA	bloqueiam	colarinho	Credenciadora	EMA/FURG	expectativa
altera	bolinho	Comercial	credibilidade	EMATER	expectativas
alternativa	bolsa	comercialização	crédito	embarcações	exponham
alto-mar	bolso	comerciante	credores	emergência	exportação
aluguel	bombas	Comerciários	crescimento	emissão	extras
alumínio	Boutique	comércio	CRISE	Empreendimentos	EXTRATO
Ambiente	brócolis	comida	crueldade	empregada	extratos
amônia	buffet	comidinha	CTPS	empregado	fábrica
andaimes	caçambheiro	comissário	cuiá	empregador	falência
animação	cadeiras	comitiva	culpa	empregadores	falha
antecipadamente	CAIXA	complementação	cultivo	empregados	falhas
aporador	caixas	compostos	custo	empregos	farinha
aposentadoria	calabresa	Compra	débito	empresários	farmácia
apressar	calma	COMPRANDO	décimo	empresas	fatura
APROVEITEM	camarão	compromisso	decisão	empréstimos	feira
Aquacultura	caminhão	Conciliação	dedicação	encanador	feirinha
aquaviários	caminhos	condeiteiro	deferimento	encomenda	Fenaban
Ar-condicionado	camomila	confia	degustar	encomendar	Ferramentas
Arena	campanheiros	Confiança	Delivery	encomende	ferreiro
argumento	campeiro	conformidades	demissão	ENERGIA	FGTS
armador	Cancelamento	conscientização	Demissões	ensino-aprendizagem	fidelidade

Tabela G.2: Termos Representativos para o aspecto Economia. Parte 2.

Fila	indenização	Micro-ondas	percepção	projetos	restaurantes
financiamento	indeterminado	milho	perfumaria	PROMETEREM	resultados
financiamentos	indumentárias	moído	perito	promoções	retribuição
fiscal	Indústria	Molhos	PERSONALIZADO	prontuários	retroescavadeira
foco	indústrias	MONTADOR	perspectiva	Pronúncia	reuniões
folhas	infinito	moradia	perspectivas	proprietário	REVENDEDORA
fome	Ingredientes	motivo	pesca	proteína	revogatório
Força	Ingresso	motoniveladora	pescador	providências	riqueza
força-tarefa	iniciante	Movimentação	pescadores	publicação	risco
formigas	inox	municípios	pescados	PUBLICAÇÕES	RODÍZIO
fornecedores	inteiros	navios	pesquisa	publicidade	roteiro
fornecimento	INTERNALIZAR	negociações	petroleiro	purês	roubo
fragmento	investidor	níveis	Petroleiros	qualificação	Safra
fragments	investidores	nível	pilchas	quantia	saídas
FRANQUEADO	Investimento	norma	PIONEIRO	quantidade	salada
franquia	iscas	novidades	pirataria	quebra	salariais
freezers	isenção	nutricionista	piso	Questiona	salário
frutas	itens	oferecendo	planilha	químico	sapataria
funcionamento	janta	ofertas	plataforma	ração	saque
funcionario	jantar	óleo	plataformas	receita	secadora
fundamento	Jornada	olho	POBRE	receitas	segmento
fundíarias	Juros	opera	POLÍTICOS	recuperação	Segurança
GANHOS	lanche	operador	pólo	recurso	Seguranças
Garçons	lancheria	Operadora	ponto	REDECARD	seguro-desemprego
gastronomia	LANCHES	operadores	Porção	redução	SEMPRE
gatos	lanchonete	oportunidade	porções	reestruturação	serralheiro
geladeira	laudo	Orçamentárias	porteiro	referendo	serviço
Geleias	Lava	orçamentos	portos	Reformas	serviços
geração	lazer	órgão	pós-paga	refrigeração	Shopping
Gerenciais	leilão	ouro	pós-pago	refrigerante	show
gerente	Leite	ovos	pós-venda	regra	SindiBancários
gigante	licenciamentos	P-58	possibilidades	reinserção	sindicalistas
golpe	liminar	P-74	pratos	reivindica	Sindicato
Governo	limitação	P-75	PRÉ-FABRICADAS	reivindicação	sindicatos
Gramado	locução	P-77	Pré-Sal	reivindicações	Sindilojas
grevistas	loja	pacote	preciso	Reivindicam	SINE
guindastes	lubrificador	Padaria	Preço	relações	SIRQ
hamburguer	LUCRO	PAGA	PREÇOS	relatório	situação
hectare	maçariqueiro	pagamento	presidentes	remuneração	SMDP
Hipercard	Macarrão	pagar	Prestadora	renda	sobremesa
Hora-extra	Macarronada	PALACIO	princípios	repassam	sobrevivência
horizontes	Manifestação	parceiro	probabilidade	reposição	soldador
Hotel	marcas	parcela	Processos	repositor	solidariedade
idiomas	MARGENS	partes	produção	requerimento	solidarizo
Igressos	marmitas	participantes	produtividade	Rescisão	solos
ilegalidade	Marmitex	Patrão	produtores	Reservas	soluções
imitações	Master	patrões	professor	responsável	sucos
impacto	matérias	Patronagem	profissão	ressalto	suflê
impasse	MERCADO	Peça	profissionalismo	ressocialização	supermercado
implantação	mercadorias	pendências	Programa	restaurante	supervisor

Tabela G.3: Termos Representativos para o aspecto Economia. Parte 3.

suprimento	tensão	trabalhos	Valor	veneno
sushi	tolerância	tradicionalistas	Valores	ventilação
suspensão	TOMATE	transporte	valorização	verbas
tábuas	tonelada	treinamento	varredor/capinador	Viabilidade
tanque	toneladas	TREINO	vazamentos	vigente
tanques	total	tristeza	veiculação	viveiros
tarifa	trabalha	troca	venda	voltam
taxa	TRABALHADOR	Troco	vendas	Whats
táxi	trabalhadoras	vaga	vendedor	WORKSHOP
tele-entrega	trabalhadores	VAGABUNDOS	vendedora	zooctenista
teles	trabalhem	vagas	vender	Zoológico

Tabela G.4: Termos Representativos para o aspecto Educação

adeptos	Co-orientador	Dreads	IFRS/Campus	PIBITI/IFRS/CNPq	tenores
adversários	Colegial	elasticidade	Igrejinha	plástica	tentativas
âmbitos	Comédias	emparelhou	Ilhescas	poetas	times
Apatuf	comedouros	empatia	IMPA	pontaria	Título
apreciação	conclusões	encampado	incidente	portais	torcida
ARQUITETURA	conteúdos	ênfases	incursões	Pré-Cálculo	Toxicologia
arte-educadores	converte	escanteio	JOGO-TREINO	Prenda	tradutora
ArtEstação	Cosmologia	Estudantil	lotação	pretensão	transitam
ASKAB	costumes	examinadora	Manicures	Quintanilha	tratamentos
Astrofísica	CPIJ	expressões	MASP	racistas	trave
auditório	curador	Fapergs	medalhista	recepções	Tributo
avalanche	curiosidades	fibra	minicursos	regiões	tricampeão
bolsita	decreta-se	Fisioterapeutas	MONITORIAS	RGsurf	tricolor
Cadete	descobertas	força-tarefa	MPCT	ritual	turmas
caracterizações	desperdice	Fototeca	Núcleos	Saudações	USP-SC
CIBER-FÍSICOS	determinações	ganhos	ofertando	saudosa	USP/POLI
CiIDEC-Sul	diputados	gerenciador	parabenizam	séculos	validade
Cinquentenário	Dissertações	gostos	pastéis	semestrais	visibilidade
co-autora	divirta-se	IFRS	periféricos	Sict	volante/zagueiro

Tabela G.5: Termos Representativos para o aspecto Empreendedorismo.

ações	Celulose	diretora	feirantes	moderação	Renda
administração	cenário	diversidade	feiras	negociações	representante
Advocacia	Co.place	edições	Feniartes	negócios	resultados
Agenciamentos	comercializar	empresarial	filés	operações	sanitário
agenda	comerciantes	energéticas	FURG	órgãos	SEBRAE/RS
Agricultura	Comércio	Energia	futuro	palestrantes	Seminário
alta	Companhia	engenheiro	Gestão	palestras	Serviços
alternativa	compliance	eólica	importância	paralelos	setores
alvará	Conferência	esforços	incerteza	parceria	Shopping
alvarás	consumidores	estande	indústria	parcerias	situação
Âmbito	contexto	estandes	indústrias	PARTAGE	soluções
aplicações	contrato	estrutura	Inovação	patrocinadores	sucesso
apoio	Cooperativas	etnia	inovadores	peixes	Supply
apostou	coordenador	evento	instituição	perspectivas	tainha
ARBRA	coordenadora	eventos	instituições	Pesca	Tecnologia
Arranjo	coworker	expectativa	INTECOOP	pescado	Tecnologias
atividades	crise	exposição	licença	pescados	tema
balanço	debates	exposições	lonas	políticas	venda
bancas	desafio	expositor	manipulação	Polo	
bilionários	desafios	Expositores	Marítimos	potencial	
bolsa	descumprimento	Fazenda	Mercado	programação	
busca	dificuldades	fearg	Metro	projetos	
Caixa	Diretor	feira	mínimas	promete	

Tabela G.6: Termos Representativos para o aspecto Energia.

abastecimento	construção	energética	incentivos	oportunizados	secretário
acessórias	construir	EQUIPAMENTOS	indústrias	perspectivas	segredo
ADEQUAÇÃO	Consumidores	esboço	iniciem	pico	serviços
água	contratada	ESTRUTURAS	Inovação	píer	subestações
atração	contrato	fábrica	instalação	píeres	tópico
Bolognesi	desligamento	Fepam	licença	POSTE	transmissão
caldeiras	desligamentos	Fonte	licenças	potência	Triunfo
capacidade	distribuição	funcionamento	licenciamento	presenças	turbina
Captação	dólar	Furg	Líquidos	RECONDUTORAMENTO	turbinas
ciclo	eficiência	gasoduto	meia	REDE	Usina
Companhia	Efluentes	geração	Nema	regaseificação	vapor

Tabela G.7: Termos Representativos para o aspecto Governança. Parte 1.

ABALADAS	andamento	bancário	cidadão	construções	demandas
ABALADO	Anotações	bancários	cidadãos	construindo	democracia
Abono	antepenúltima	barcaças	cidades	construtoras	DENUNCIA
abstenção	anulação	bases	Cinzas	consultor	denúncias
abstenções	anúncio	BASTA	cirurgia	contábil	Deputados
Aceleração	anúncios	benefício	CNPJ	contabilidade	desabastecidos
acesse	APAE	bens	CNTE	contadora	Desafios
Acessibilidade	Apátridas	bilionário	COAÇÃO	contas	desajuste
acesso	apelo	boicote	cobertores	contradições	desemprego
acionamento	aponta	BOMBA	cofres	contrapartida	desigualdade
ações	apontam	brasileiros	coibição	CONTRATAÇÕES	despesas
acomodação	aposentadorias	Brigada	coligações	contrato	detentos
acordos	aprovação	BURACO	coligados	contratos	determinação
acusação	apuração	cadastramento	combater	contribuição	DEVEDORES
acusações	AQUICULTURA	cadastros	COMDEC	contribuintes	diferença
adequações	aquisição	cadelas	comercialização	controle	diferencia
adesão	argumento	cafeteria	comícios	contrução	dificuldades
adjunto	argumentos	calamidade	Comirat-RG	convenções	DIGNIDADE
administração	arrecadação	calcário	comitê	convênio	diminuições
administrações	articulação	calendário	Comitiva	convênios	dinâmica
adquirir	Asdep	CAMBADA	comodidade	Coordenação	dique
Advocacia	asilo	campanhas	Comparecimento	coordenador	diretores
advogada	ASSEMBLEIA	Canais	COMPDEC	coordenadora	diretriz
ADVOGADOS	assembleias	candidata	COMPDEC/RG	Coordenadoria	diretrizes
aerogeradores	assinatura	candidatura	competência	CORPORATIVISTA	discussão
afirmação	assinaturas	candidaturas	complementação	cortada	discussões
afirmações	Assistência	canteiros	complexidade	cortar	dívidas
afirmar	atacam	Capela	complexidades	CPERS	doadora
Agasalho	atendimento	Capital	componentes	creches	doadoras
agasalhos	atingem	caput	composição	crise	documento
agenda	atitude	cargo	compras	Critério	documentos
agilidade	atitudes	cartazes	comprometem	CRLV	dólares
Agradecimento	atos	CARTEL	compromisso	cronograma	doleiro
agravamento	atraso	Cartório	compromissos	Cunha	ecocardiogramas
agressões	atrasos	cartórios	Comprovante	CUSTARÃO	ecodopplers
agricultores	atributos	casco	Comunidade	custo	ecografias
agricultura	atuação	cassação	concessão	custos	economista
AHTS	Audiência	castrações	conclusão	D'agostini	ECOSUL
alinhar	auditoria	CCMar	Concorrência	dados	edital
ALTERNATIVAS	ausência	CDES	concorrentes	DATC	efeitos
alunos	Autárquico	CDES-RG	CONCURSOS	debates	elaboração
Ambiente	automação	CELERIDADE	condicionado	débitos	eleição
ambientoterapia	autonomia	cemitério	configurações	decisões	ELEIÇÕES-2014
âmbito	autoridades	cenas	conluio	declaração	eleitor
ambulância	auxílio	chalé	conquista	Decreto	Eleitorado
ambulatoriais	avançar	chalés	Conselho	DEFINIÇÃO	eleitorais
ameaças	bacias	chapa	Conselhos	delação	eleitores
amigos	bairro	chefe	consequências	deliberações	embasamento
ampliação	bairros	chocolataria	consolidação	demandas	EMBRIAGUES
andam	balneário	Cidadania	construção	demandam	emergência

Tabela G.8: Termos Representativos para o aspecto Governança. Parte 2.

empenho	envelopes	Escritórios	MEC/FNDE
emplacamento	envolvimento	escrivão	morada
empreendimentos	equipamentos	esfera	Neoliberalismo
empreiteiras	equipes	esforço	pedaços
empresariais	Erradicação	esgoto	ressonâncias
empresarial	escalão	esquema	siderurgia
empréstimo	escândalos	estações	sucessória
encaminhando	esclarecimentos	estupidez	tentáculos
endividamento	escolares	graminha	TOLERÁ
engajamento	escolaridade	hoteleiro	vergonhas
entidade	escolas	Idealismo	xingam

Tabela G.9: Termos Representativos para o aspecto Meio Ambiente. Parte 1.

#MataAtlântica	belezas	custou	Fazenda	lixo	Oceanografia
#recicleseushábitos	Biodiversidade	d'água	fenômeno	lucro	oceano
#reduzaconsumodeplásticos	biomagnificação	danifica	Fepam	madeiramento	oeste
abelhas	biomas	danos	firme	mamíferos	óleo
abriga	botos	decibelímetros	flora	manejo	Ondas
abrigue	busquem	declarações	Floresta	mapas	operação
Abundância	Cachoeira	degradação	Florestal	máquinas	operações
ações	calor	denúncias	fontes	maranduvás	organismos
acúmulo	camarão	departamentos	FOPEMEPE/RS	maré	painel
advento	caramujo	depósito	fornecedor	marés	paisagem
afunda	carências	depósitos	fortalecimento	margem	Pancadas
afundar	casca	desafio	fragmentação	margens	PAPELAO
agricultura	cattleyanum	descargas	fragmentos	marinhos	parcerias
Agrônomo	caule/tronco	deslizamentos	FURG	marítima	partículas
água	cavalos	desmoronamentos	gama	material	Patos
águas	cenários	despejo	gases	máxima	Patram
alagamentos	centros	desperdício	gastrópode	máximas	Patrimônio
alugamento	chove	destine	Genoves	MÉDIA	PDAU
alturas	chover	detalhe	golfinhos	mediação	peixes
amanhecer	chuva	diálogos	Governos	meteorológicas	peltophoroides
Ambientalistas	chuvas	dimensão	grama	Meteorologista	perigo
ambiente	chuvisqueiro	dispersão	granizo	métodos	pescaria
ambientes	ciclo	dragas	Habitat	microesférulas	pescado
anfíbios	CICLONE	Ecólogo	haste	micropartículas	pesquisa
animais	cientistas	ECOSERVICE	hídricos	Microplásticos	pesquisadores
anormalidades	cifra	ecossistema	Hotel	milhas	picos
aparelhos	clima	ecossistemas	Impacto	milho	plantas
aparições	climáticas	ecossistêmicos	Impactos	mimosaefolia	plásticas
APPs	coberturas	embarcação	incêndio	mm/dia	plástico
aproveitamento	cobra	embargo	incidência	mm/h	plásticos
aquaplanar	cobras	emissão	indicadores	modificações	pluma
Aquicultura	coincide	enchente	insetos	Molhes	plumas
Arborização	coincidencia	enchantes	instabilidade	Monumento	polêmica
Área	comportamento	encosta	intensidade	mosquito	polemizar
areia	composição	encostas	inundações	mosquitos	poluentes
Arquipélagos	concentração	entorno	Irrigação	muda	populacionais
arrecada	concha	equilíbrio	isoladamente	mudas	praias
árvores	Conclusão	erosão	jacaré	nações	precaução
atenuamento	condições	espécie	kilometros	Nascente	preservacao
atmosfera	conscientizaçao	espécies	lagartas	natureza	previsão
autorização	consequências	espiga	lagartixa-das-dunas	navegação	Previsões
aves	conservação	Estação	Lagoa	nebulosidade	procedimentos
avifauna	consumo	estragos	Laguna	neve	processos
baías	continentes	estuário	lama	névoa	projeções
bairro	Controle	estuários	latas	ninhos	Proliferação
baixas	Cooperação	estudo	LCBC	níveis	promete
Barco	cordão	exploração	licença	normas	proporções
barro	corujas	extinção	Licenciamento	nutrionistas	Proteção
beira	cultivo	Extrativista	limpeza	nuvens	Psidium
beleza	cultivos	fauna	liso	observações	pulverizar

Tabela G.10: Termos Representativos para o aspecto Meio Ambiente. Parte 2.

químicos	remoção	romanzoffiana	situações	território	vegetação
raios	reprodução	ruído	SMMA	tesouro	vegetais
rajadas	répteis	ruídos	SNUC	transbordamentos	vegetal
rastilhão	réptil	sacola	solta	transtornos	velocidades
ratos	Reserva	sacolas	sonora	tratamento	ventania
recolhimento	residuos	sacos	sudoeste	trator	VENTO
recuperação	resistência	Safra	superfície	trovoada	Ventos
redes	respeito	SEARG	suspenção	trovoadas	veraneio
redução	responsabilidade	secretário	sustento	tubinhos	veterinário
refeição	ressaca	sedimentação	tainha	tuco-tuco	viagens
refeições	Ressacão	sedimento	taturanas	UMIDADE	vias
Região	ressacas	sedimentos	temperatura	urbanização	virar
regiões	resto	SENSAÇÃO	temperaturas	Vales	vizinhança
regulação	rios	serração	tempestades	Vanádio	
relatório	riscos	Situação	terebinthifolius	variação	

Tabela G.11: Termos Representativos para o aspecto Mobilidade.

acessibilidade	carga	Detran-RS	lombada	pico	Sindicam
acesso	cargas	dividirá	lombadas	piloto	situação
accidentalidade	carona	duplicação	madrugada	Pinto	SMMUA
acidentes	carreta	ERS-734	manifestantes	placa	subgrupo
ações	carretas	espalharam	manobrar	placas	subgrupos
acostamento	carro	estação	manobras	PLANTÃO	sufoco
adequação	carro/carro	estacionamento	MAPEIA	policiais	tarifa
adubo	Carro/moto	estada	máquinas	Prefixo	tarifárias
advogado	CARROÇA	Estradas	margens	prevenção	tarifas
ajuda	carroceiro	estudantes	Masiero	Profissionalização	táxi
ajude	CARROS	exceção	migração	proibir	terminais
ajustes	Cassino/Rio	Fabril	Mobilidade	Pronto-Socorro	Trabalhador
AMBULÂNCIA	cavalo	FERE	moto	publicitárias	trabalhistas
armazenamento	cena	ferimentos	motocicleta	quilômetro	trafegam
aterro	choque	Fiscal	motocicletas	quilômetros	trafegar
atraso	ciclista	fiscalização	motociclistas	reconsideração	tráfego
atravessar	ciclistas	fluxo	motofretistas	reformulações	tranqueira
AUTO	ciclístico	fomentar	Motor-Casa	registrar	trânsito
autuação	ciclovia	fratura	Motorista	regulamentação	transportadoras
auxiliares	cinto	fraturas	MOTORISTAS	Renovações	Transportadores
AV.Pelotas	circulação	frota	motos	RESPEITA	transporte
avenida	colisão	frotas	mototaxistas	respeitarem	transportes
avenidas	colisões	Furg	NAVIO	retorno	Transposul
Barra/Trevo	composto	gratuidade	óbito	retrato	trechos
barranco	compromete	gratuidades	obras	retroportuários	trem
BATE	concentram	guarda	ocorrência	Riograndinos	tributos
bicicleta	Condutor	hipóteses	ônibus	rodovia	TRIPULANTE
bloqueia	condutores	iluminação	operações	rodoviária	UPF-RS
BR-116	congestionamento	Imprudência	orientações	ROTULA	usuários
BR-392	controle	incremento	página	RS-040	veículo
BR-471	correção	indexador	paradas	RS-118	veículos
calçada	Cotista	indignação	Paralisação	RS-239	velocidade
calçadas	críticos	infração	param	ruas	velocímetro
cálculo	cruzamentos	Infraestrutura	passageiro	secretário	verifica
caminhão	cuidado	instalação	passageiros	semáforo	viaduto
caminhões	curva	instantâneo	passe	sensibilização	viaja
caminhoneiro	DAER	isenção	passeios	sequelas	vias
caminhoneiros	danificando	lados	patronais	sextas-feiras	vidas
camionete	danos	lancha	pedestres	sinal	vítima
caos	deslocamento	lesões	perímetro	sinalização	vítimas

Tabela G.12: Termos Representativos para o aspecto Saúde.

abandono	carnes	doutor	hospital	Multivacinação	reforma
ações	caxumba	Drogas	hospitalar	Mycobacterium	Rosa
acondroplasia	CEANNE	DSTs	Hospitalares	neoplasias	sangue
aditamento	certificação	ecografia	hospitalizações	Neurologia	Santas
adolescentes	cigarro	Educação	HU-Furg	Neuros	secretário
Aedes	cirurgia	eletrocardiograma	HU-Furg/Ebserh	nicotina	secretário-adjunto
afrodescendentes	cirurgias	eletrocardiogramas	HU/Furg	nutricionista	segundo-secretário
agentes	clínicas	Enfermeiras	imunização	obesidade	serviço
agravamentos	coagulase	Enfermeiros	Imunizações	óbito	servidores
aids	coliformes	equipamentos	incidência	observa	sifilis
alcool	coloniais	Equipes	infecções	Oftalmologia	Sindisaúde
alergia	combate	especialistas	influenza	órgãos	sintoma
alimentação	combater	esportes	ingestão	paciente	sintomas
alunos	complicações	estabelecimentos	inspeção	pacientes	situação
Alzheimer	compromiço	estética	integralidade	palestras	SMCAS
ambulâncias	condições	estoque	integridade	palhaços	subtipos
ameaça	Conselho	estoques	internações	Papilomavírus	suplementos
amostras	contagens	exame	internados	paredes	terminar
ampliação	contaminação	exames	irresponsabilidade	parto	tipo
Animal	contrato	expansão	IST/AIDS	penitenciárias	titulares
antecedencia	Coordenadoria	ferimento	IST/Aids/Hepatites	perspectivas	transporte
antibióticos	corpo	ferimentos	ISTs	PESQUISA	transmissão
apoio	Cosems	físicos	laboratorial	PÍLULA	tratamento
aponta	CosemsRS	foco	laboratórios	plantonista	triglicerídeos
apontam	cuidados	fosfoetanolamina	Lacen	pneumonias	tuberculose
apresenta	cuidando	fragilidade	leitos	portadores	tuberculosis
asplenia	cumulo	fumantes	leve	portarias	tumor
assistênciA	cura	FURG	liberdade	posto	tumores
ATENDIMENTO	danos	genética	longevidade	preservativos	UBSF
ATENDIMENTOS	demandas	geneticista	mama	pressão	UFPel
atrasa	demora	gestação	mamas	prevencao	unidade
bactéria	dependênciA	gestantes	mamografia	prevista	unidades
bairro	Diabetes	ginecologista	Materno	prioritários	urologista
beneficiários	diagnóstico	glicemia	médica	procedênciA	vacina
biologia	diagnósticos	graças	medicação	Pronto-socorro	vacinação
Bipartite	dirigente	gripe	medicações	próstata	vacinações
boca	distribuição	guloseimas	medicamentos	proteção	vacinas
breve	doação	h1n1	Medicina	pública	veterinário
caderneta	doações	hábitos	médico	público-alvo	Vigilância
calamidade	doadores	higiene	menores	puérperas	vírus
campanha	doença	hipertensão	metodologias	qualidade	visem
câncer	doenças	hipocondroplasia	micobactéria	quimioterapia	vistoria
carcerário	dosagem	HIV/AIDS	microbiologia	raiva	Vítima
cardiologia	dose	hormonais	mortalidade	Recém-nascidos	
cardiologista	doses	hospitais	mosquito	rede	

Tabela G.13: Termos Representativos para o aspecto Segurança.

#ASSALTADO	calma	conjunção	espuma	moto/carro	repuxo
ação	calmo	constata	estocada	obstáculos	reservista
acertaram	carcerário	contraprova	executor	oferecem	rifle
advertências	carro/carro	cucuruto	êxtase	pano	ROCAM
agente	casacos	decorrência	Faroeste	peito	salento
agressões	casarões	delegada	fatalidade	perfurações	SEQUESTRA
aperfeiçoamento	catador	delito	flanelinha	PLANTÃO	SEQUESTRAVA
aquaplanou	causadores	desclassificação	formações	Plaza	skatistas
ARREMESSAM	centímetros	desconhecimento	frase	Pneus	táxis
ASSALTAVAM	Centro/Barra	destalhes	frentistas	porte	tecido
assusta	cerveja	digestão	FUMO	pranchas	tijolada
ateando	chão	discussão	Furg/Carreiros	pretensão	tijolinhos
balneários	chegará	disparo	furto/abigeato	PÚRPURA	transporte
banhista	chuveiro	efetividade	Furto/descuido	quadrante	vagina
barba	cicatrizes	efetua	idades	quadril	valas
batidas	colchões	envenenamento	incêncio	rabetas	veneno
bonde	colocam	envólucros	indiciamento	racismo	veranista
brigadianas	combustível	escoriação	indigestão	redondezas	violação
bueiro	comerciária	ESFAQUEAMENTO	inexistência	reduito	
cachorros	companheiro	ESPANCANDO	inexistentes	refrigerantes	
calçou	confrontos	esportistas	justiceiros	registram	

Tabela G.14: Termos Representativos para o aspecto Urbanismo.

#CidadeLimpia	balanços	estufas	investimento	proximidade	SMHARF
acabamentos	BALNEÁRIO	facilita	lazer	psicólogos	SMHRF
ação	Barra	ferragens	Lixão	quadra	suportes
Aceleração	brinquedo	floreiras	meio-fio	quartos	sustentação
Acessibilidade	brinquedos	Fundiária	meios-fios	realinhamento	terraplanagem
acesso	calçamento	FURG	melhorias	reestruturação	terreno
acessos	Campanha	galpão	monitores	Regularização	térreo
ações	ciclístico	gangorras	obras	remoção	transição
acomodação	colaboração	grama	organização	reparo	transtornos
administração	construção	Habitação	pintura	reparos	travessa
água	construindo	habitacionais	pisos	resíduos	tubulação
alvenaria	cortaram	habitacional	pistas	retenção	tubulações
animais	corte	habitações	planejar	retirada	turnos
aparelhos	CRAS	harmonia	Playground	Rincão	uniformidade
apartamento	cuida	harmônica	pluviais	roedores	urbanização
apartamentos	danificação	harmonização	poda	rótula	viabilidade
apropriação	depredação	higienização	postes	ruas	vias
Arquipelágos	Dragões	iluminação	prejudicam	Ruffier	wolff
árvores	drenagem	Infraestrutura	programando	Saneamento	zelo
assistentes	educadores	inserção	projeta	Saraiva	
autorização	entorno	instala	projetos	secretário	
avenida	equipes	instalações	propagação	Senandes	
avenidas	escoamento	integração	PRORUAS	Serviços	

APÊNDICE H – TERMOS REPRESENTATIVOS MAIS FREQUENTES A PARTIR DE TODOS DOCUMENTOS CLASSIFICADOS



Figura H.1: Nuvem de palavras para o aspecto Economia.



Figura H.2: Nuvem de palavras para o aspecto Educação.



Figura H.3: Nuvem de palavras para o aspecto Empreendedorismo.



Figura H.4: Nuvem de palavras para o aspecto Energia.



Figura H.5: Nuvem de palavras para o aspecto Governança.



Figura H.6: Nuvem de palavras para o aspecto Meio Ambiente.



Figura H.7: Nuvem de palavras para o aspecto Mobilidade.

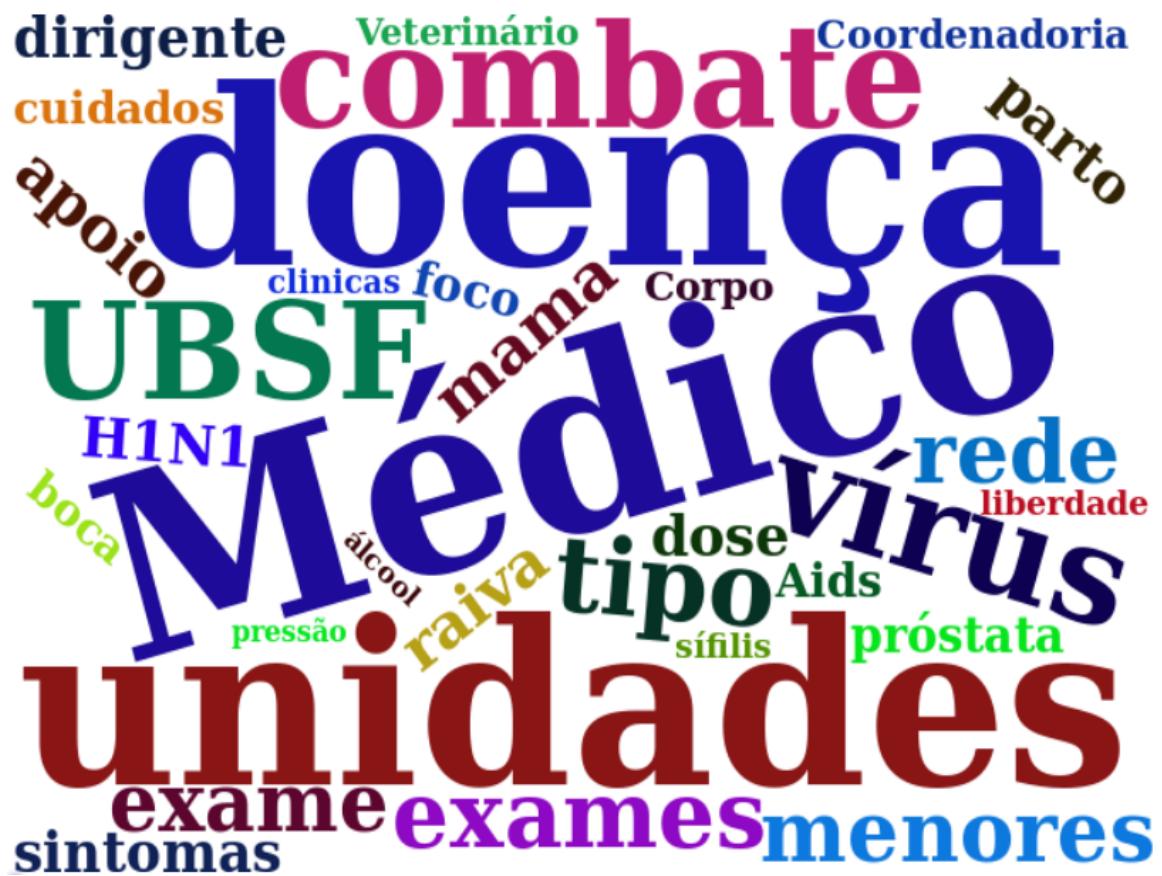


Figura H.8: Nuvem de palavras para o aspecto Saúde.



Figura H.9: Nuvem de palavras para o aspecto Urbanismo.

ANEXO A – TABELAS DE INDICADORES DE CIDADES INTELIGENTES

Tabela A.1: Indicadores de Cidade Inteligente propostos pela *Urban Systems*. Parte 1.

Tabela A.2: Indicadores de Cidade Inteligente propostos pela *Urban Systems*. Parte 2.

	INDICADOR	Mobilidade e Acessibilidade	Urbanismo	Energia	Tecnologia & Inovação	Saúde	Segurança	Educação	Empreendedorismo	Governança	Economia
Saúde	Leitos por Habitantes										
	Leitos de Internação (UTI e Semi)					X					
	Médicos por habitantes					X					
	Cobertura populacional da Equipe de Saúde da Família					X					X
Segurança	Número de conluiantes no setor de saúde					X					
	Homicídios					X					
	Acidentes de Trânsito	X				X					
	Policiais, Guarda-civis Municipais e Agentes de Trânsito					X					
Educação	Matrícula escolar na rede pública online					X					
	Vagas em Universidade Pública					X					
	Nota Enem					X					
	Docentes com Ensino Superior					X					
Empreendedorismo	IDEB - Anos Finais					X					
	Hora-aula diária média					X					
	Novas empresas de tecnologia					X					
	Polos Tecnológicos					X					
Governança	Crescimento Empresas de Economia Criativa					X					
	Incubadoras					X					
	Micro Empresas Individuais - MEI					X					
	Sebrae					X					X
Economia	Escolandade do Prefeito					X					
	Prefeitura com Site na Internet, serviços e notícias					X					
	Indice Firjan					X					
	Despesa Municipal com Segurança					X					
	Despesa Municipal com Saúde					X					
	Despesa Municipal com Educação					X					
	EBI Escala Brasil Transparente					X					
	Conselhos Municipais					X					
	PIB per Capita					X					
	Renda Média dos Trabalhadores					X					
	Crescimento Empresarial					X					
	Crescimento Empregos Formais					X					
	Empregos Independentes do Setor Público					X					
	Empregabilidade					X					
	Receitas não oriundas de Transferências					X					
						X					