



**UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO**  
**CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

**Mídias são parciais! Será? Uma ferramenta de análise automática dos canais de notícia em relação às figuras políticas do Brasil**

LUIZ ORLANDO PEREIRA BRAGA

RECIFE

2017

LUIZ ORLANDO PEREIRA BRAGA

**Mídias são parciais! Será? Uma ferramenta de análise automática dos canais de notícia em relação às figuras políticas do Brasil**

Monografia apresentada ao Curso de Bacharelado em Ciência da Computação da Universidade Federal Rural de Pernambuco, como requisito parcial para obtenção do título de Bacharel em Ciência da Computação.

Orientador: Marcelo Marinho

Coorientador: Rafael Ferreira



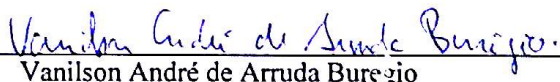
MINISTÉRIO DA EDUCAÇÃO E DO DESPORTO  
UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO (UFRPE)  
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO


<http://www.bcc.ufrpe.br>

**FICHA DE APROVAÇÃO DO TRABALHO DE CONCLUSÃO DE CURSO**

Trabalho defendido por Luiz Orlando Pereira Braga como requisito para conclusão do curso de Bacharelado em Ciência da Computação da Universidade Federal Rural de Pernambuco, intitulado *Mídias são parciais! Será? Uma ferramenta de análise automática dos canais de notícias em relação às figuras políticas do Brasil*, orientado por Marcelo Luiz Monteiro Marinho e aprovado pela seguinte banca examinadora:

  
\_\_\_\_\_  
Marcelo Luiz Monteiro Marinho  
DEINFO/UFRPE

  
\_\_\_\_\_  
Vanilson André de Arruda Buregio  
DEINFO/UFRPE

  
\_\_\_\_\_  
Suzana Cândido de Barros Sampaio  
DEINFO/UFRPE

# Agradecimentos

Agradeço primeiramente aos meus pais, Luiz Manoel Braga da Costa e Luziana Maria Pereira Braga, que procuraram superar todos os obstáculos para que pudessem me dar uma melhor educação e que sempre me guiaram para o caminho do bem. Agradeço aos meus irmãos e irmã pelo companheirismo e cobranças que ajudaram na minha formação como pessoa. À Carina Sá Leitão pela amizade sem igual e por toda a paciência que tem de me ouvir e me dar conselhos, e me estabiliza nos momentos onde os problemas parecem não ter fim. Agradeço ao meu orientador Marcelo Marinho, pelo conhecimento e oportunidade de me aceitar como aluno. Uma pessoa no qual devo me inspirar tanto profissionalmente quanto como pessoa. Agradeço também à Marcelo Lacerda por me fornecer conhecimento e me guiar nesse trabalho. Finalmente, agradeço também aos meus companheiros leais, Cristovam Segundo, Evellinne Maria e Amaro Virgínio, que são grandes amigos e pessoas bastante inteligentes, onde tive o prazer em conhecê-los.

# Resumo

Atualmente existe uma grande quantidade de notícias sendo publicadas diariamente, tornando um grande volume de dados. Sendo assim, pensando na análise desses dados de uma maneira questionadora, para determinar o que os diferentes portais de notícias estão querendo relatar, foi proposto uma ferramenta para realizar a análise sentimental do texto das notícias. Essa ferramenta realiza métodos de inteligência artificial e tratamento de texto para determinar a polaridade geral das notícias sobre determinado envolvido. Esse estudo foi focado em políticos por se tratarem de figuras públicas e que precisam de uma fiscalização e análise aprofundada. Em suma, a ferramenta possui como objetivo realizar uma visualização de qual sentimento os canais de notícias querem mostrar quando estão relatando sobre determinado político, fazendo com que seja possível identificar a parcialidade dos mesmos.

**Palavras-chave:** Inteligência Artificial; análise de sentimento; categorização;

# Abstract

In recent days, there is a great quantity of news being published daily, making a great volume of data. Thus, keeping in mind the data analysis in a questionable way to determine what the different newspapers are really trying to report, it has been the purpose of a software that will perform the sentiment analysis of the text given by the news. This software uses artificial intelligence and text processing tools to determine the polarity of the news about distinct individuals. The study was focused on Brazilian politicians due to being public figures and the need for a supervised and in-depth analysis. In short, the software tool has as its objective the visualization of the sentiments the different newspapers have about specific politicians, this way will be possible to identify the partiality of them. At the end, it has been possible to notice how the newspapers behave when talking about those politicians.

**Keywords:** Artificial intelligence; sentiment analysis; categorization;

# Lista de ilustrações

Figura 1 – Etapas do processamento de texto . . . . .	19
Figura 2 – Arquitetura do Scrapy (Scrapy developers, 2016) . . . . .	22
Figura 3 – Método de pesquisa utilizado no trabalho . . . . .	27
Figura 4 – Modelo de mineração de dados inspirado em Aranha (2007) . . . . .	27
Figura 5 – Exemplo de notícia armazenada no banco de dados . . . . .	29
Figura 6 – Exemplo de conjunto de palavras de características . . . . .	31
Figura 7 – Arquitetura da Ferramenta Proposta . . . . .	36
Figura 8 – Exemplo de palavras de características de uma notícia . . . . .	41
Figura 9 – Taxa de acerto obtida através da base de teste . . . . .	41
Figura 10 – Análise de sentimento da Carta Capital sobre Dilma Rousseff . . . . .	43
Figura 11 – Análise de sentimento da Carta Capital sobre Renan Calheiros . . . . .	44
Figura 12 – Comparativo de acerto . . . . .	45

# Lista de tabelas

Tabela 1 – Critérios utilizados para classificação de notícias. . . . .	30
Tabela 2 – Captura total de notícias . . . . .	38
Tabela 3 – Taxa de acerto dos classificadores . . . . .	41
Tabela 4 – Quantidade de notícias em quatro meses . . . . .	42
Tabela 5 – Tabela de notícias negativas . . . . .	54
Tabela 6 – Tabela de notícias positivas . . . . .	64
Tabela 7 – Tabela de notícias neutras . . . . .	71



# Lista de abreviaturas e siglas

URL	<i>Uniform Resource Locator</i>
JSON	<i>JavaScript Object Notation</i>
BSON	<i>Binary JavaScript Object Notation</i>
NLTK	<i>Natural Language Toolkit</i>

# Sumário

	<b>Lista de ilustrações</b>	<b>6</b>
<b>1</b>	<b>INTRODUÇÃO</b>	<b>11</b>
1.1	Motivação	11
1.2	Justificativa	12
1.3	Objetivos	13
1.3.1	Objetivo Geral	13
1.3.2	Objetivos Específicos	13
1.4	Estrutura do trabalho	13
<b>2</b>	<b>REFERENCIAL BIBLIOGRÁFICO</b>	<b>15</b>
2.1	Trabalhos Relacionados	15
2.2	REST API	16
2.3	Mineração de Textos	18
2.4	Processamento de Linguagem Natural	19
2.4.1	Análise de sentimento	20
2.5	Ferramentas	21
2.5.1	Scrapy	21
2.5.1.1	Arquitetura	21
2.5.2	MongoDB	24
2.5.3	Python NLTK	25
<b>3</b>	<b>METODOLOGIA</b>	<b>26</b>
3.1	Método de pesquisa	26
3.2	Determinação do modelo para mineração de texto	27
3.2.1	Coleta de texto de notícias	28
3.2.2	Pré-processamento	29
3.2.3	Classificação dos textos	30
3.2.4	Análise de sentimento	31
<b>4</b>	<b>DESENVOLVIMENTO DA FERRAMENTA</b>	<b>33</b>

<b>4.1</b>	<b>Levantamento de Requisitos</b>	<b>33</b>
4.1.1	Requisitos Funcionais	34
4.1.2	Requisitos Não Funcionais	35
<b>4.2</b>	<b>Arquitetura</b>	<b>35</b>
<b>4.3</b>	<b>Testes Unitários</b>	<b>37</b>
<b>5</b>	<b>RESULTADOS</b>	<b>38</b>
<b>5.1</b>	<b>Dados coletados</b>	<b>38</b>
<b>5.2</b>	<b>Pré-processamento</b>	<b>39</b>
<b>5.3</b>	<b>Validação do modelo de classificação</b>	<b>40</b>
<b>5.4</b>	<b>Análise de sentimento</b>	<b>42</b>
<b>6</b>	<b>CONCLUSÃO</b>	<b>46</b>
<b>6.1</b>	<b>Trabalhos Futuros</b>	<b>47</b>
	<b>ANEXOS</b>	<b>49</b>
	<b>ANEXO A – DESCRIÇÃO DE REQUISITOS</b>	<b>50</b>
<b>A.1</b>	<b>Requisitos Funcionais</b>	<b>50</b>
A.1.1	Captura	50
A.1.2	Mineração de Texto	51
A.1.3	API	52
<b>A.2</b>	<b>Requisitos Não Funcionais</b>	<b>52</b>
A.2.1	[RNF01] Segurança dos dados	52
A.2.2	[RNF02] Desempenho	52
A.2.3	[RNF03] Disponibilidade do serviço	53
	<b>ANEXO B – NOTÍCIAS DE TESTE</b>	<b>54</b>
	<b>REFERÊNCIAS</b>	<b>74</b>

# 1 Introdução

Neste capítulo será apresentada a motivação para a realização deste trabalho e os objetivos esperados.

## 1.1 Motivação

A sociedade democrata conquistou o direito do acesso à informação, sendo esse acesso alterado com o passar dos anos. Na sociedade moderna, a utilização da mídia digital para divulgação de informações está se tornando cada vez mais popular (BARBOSA, 2015).

De acordo com *Statistical Report on Internet Development* divulgado pela *China Internet Network Information Center* (CNNIC), houve um aumento de leitores de notícias de 68.8 milhões para 461 milhões durante o período de um ano entre junho de 2012 e 2013 (CNNIC, 2013). Esse aumento foi causado pela utilização da Internet móvel, que torna possível o acesso pelos usuários em qualquer lugar e momento. Além disso, o acesso ficou mais diversificado porque vários canais de notícia estão disponíveis para visualização. No caso do Brasil, o país ocupa a quarta posição no rank de usuários de internet, com 66.4% da população utilizando a internet (CNNIC, 2016). De acordo com a Pesquisa Brasileira de Mídia (PESQUISA..., 2015), dos entrevistados que utilizam a internet, a maioria deles utilizam o *Facebook*, e a maior parte de usuários são jovens de 16 a 25 anos, que fazem parte da geração Y<sup>1</sup>. Uma outra pesquisa, realizada pelo *Pew Research Center* nos EUA (MITCHELL; GOTTFRIED; MATSA, 2015), mostrou que a geração Y visualiza mais notícias sobre política através do *Facebook* em comparação com as outras gerações (X<sup>2</sup> e *baby boomers*<sup>3</sup>). Sendo assim, é possível afirmar que a maioria dos jovens que acessam a internet prefere visualizar notícias sobre política através do *Facebook*. Dessa maneira, faz com que o *Facebook* torne-

<sup>1</sup> Pessoas que nasceram no período de 1980 até o fim dos anos 90. São conhecidos também por geração milenar.

<sup>2</sup> Pessoas que nasceram no período de 1965 a 1976.

<sup>3</sup> Pessoas que nasceram no período de 1946 e 1964.

se um meio, um caminho, para que as notícias cheguem às pessoas, reforçando a utilização de notícias publicadas digitalmente.

Existem estudos relacionados à esse problema de extrair informações de notícias na web (TICOM; SOUZA; LIMA, 2007), em que foi feita uma busca em profundidade, buscando os 5W1H (*When, What, Who, Where, Why, and How*) de forma automatizada no texto de notícias da China. Outro exemplo é uma ferramenta criada para notícias da Índia (WANJARI et al., 2014), onde foi utilizado uma busca por padrão para classificar o conteúdo das notícias. Existe também um exemplo de outra ferramenta com uma solução mais complexa (IGLESIAS et al., 2014), que envolveu utilizar classificação evolutiva com o objetivo de determinar a categoria da notícia, essa proposta consiste na criação de um módulo de regras *Fuzzy* de evolução (ANGELOV; ZHOU, 2008) com o objetivo de criar e atualizar as regras para determinar a categoria ou categorias da notícias e outro módulo com a capacidade de classificar as notícias nas categorias e agrupar os resultados mais próximos para lidar com diferentes formas de agrupamento, pois segundo o autor, as notícias não possuem uma classificação fixa.

## 1.2 Justificativa

Com a popularidade das notícias em modelo digital e um grande volume de artigos de notícia sendo publicados, a busca por informação se tornou cada vez mais necessária. Notícias são publicadas e visualizadas poucos segundos após serem lançadas, gerando assim uma grande quantidade de conteúdo todos os dias. Este conteúdo pode ser analisado e processado utilizando técnicas de mineração de dados para obtenção das informações desejadas (FELDMAN; DAGAN, 1995) (HSU, 1995), resultando assim em um problema de *Big Data*.

A definição de *Big Data*, segundo Mauro, Greco e Grimaldi (2016), de maneira simplificada, trata-se de utilizar dados em enorme volume e métodos analíticos para ser transformado em valor, trazendo padrões, tendências e associações. Valor nesse caso é definido como informação que é útil para alguém, em que foi obtida através dos dados coletados. Trazendo para o contexto do tema deste trabalho, por existir um grande volume de notícias publicadas diariamente por diversos canais de notícias, é gerado

um grande volume de dados que podem ser analisados e obter novas informações relacionados à quem as publica e sobre a quem elas são direcionadas.

Com vista a esse problema, é sugerido a criação de uma ferramenta para automatizar a classificação e identificação de notícias digitais brasileiras sobre determinado indivíduo político divulgadas pelos principais portais de notícia (G1, UOL, Folha de São Paulo, Veja, Carta Capital e Época). Buscando uma visão analítica sobre o conteúdo destes artigos de notícia e como o canal de notícia enxerga determinado político. Os portais de notícias foram escolhidos seguindo o critério de importância, credibilidade e popularidade. Existem outros canais de notícia importantes para serem estudados, mas por viabilidade e tempo, foram escolhidos apenas alguns.

## 1.3 Objetivos

### 1.3.1 Objetivo Geral

Desenvolver uma ferramenta capaz de classificar o conteúdo de notícias, realizando a análise sentimental.

### 1.3.2 Objetivos Específicos

1. Identificar na literatura um modelo de análise de sentimento e trabalhos relacionados;
2. automatizar a coleta de dados na web, através de *web mining*;
3. utilizar classificadores, como Naïve Bayes e SVM, para determinar a polaridade da notícia;
4. relacionar acontecimentos de acordo com a análise sentimental das notícias.

## 1.4 Estrutura do trabalho

Além deste capítulo introdutório, a monografia está organizada em mais cinco capítulos que consistem em Referencial Bibliográfico, Metodologia, Desenvolvimento da API e Resultados.

O capítulo 2, Referencial Bibliográfico, mostra trabalhos relacionados e descreve as tecnologias e ferramentas mais importantes para a criação da ferramenta proposta. O capítulo 3, Metodologia, descreve quais os passos tomados em cada etapa do trabalho. O capítulo 4, Desenvolvimento da ferramenta, relata os requisitos que a ferramenta irá cumprir e como a arquitetura da ferramenta foi criada, assim como a etapa de testes. O capítulo 5, Resultados, mostra os resultados obtidos com as execuções da ferramenta e quais informações foram obtidas através dos dados coletados.

## 2 Referencial Bibliográfico

Este capítulo relata trabalhos relacionados ao proposto e quais contribuições eles tiveram. Além disso, o capítulo apresenta uma descrição das principais tecnologias e bibliotecas utilizadas na ferramenta proposta.

### 2.1 Trabalhos Relacionados

O trabalho de Efthymios, Theresa e Johanna (KOULOUMPIS; WILSON; MOORE, 2011) trata-se de um estudo sobre o uso de características linguísticas para realização de análise de sentimento, como semântica, polaridade sentimental da palavra e reações dos *emojicons* utilizados, onde mostraram-se promissoras em outras formas de texto diferente da utilizada no trabalho, que são textos do *Twitter*. Durante o estudo, foram realizados experimentos utilizando textos do twitter que foram capturados através de *hashtags*. Como resultado, foi obtido que em textos do *Twitter* a utilização de características de *part-of-speech*, palavras que possuem propriedade gramatical similar, não pareceu promissor para o domínio de textos pequenos. Entretanto, a utilização de características de léxico sentimental mostrou-se promissor.

A diferença do trabalho de Kouloumpis, Wilson e Moore (2011) para os outras pesquisas similares, é o foco em dados *microblogging*, que são geralmente textos pequenos como *tweets* e obtenção de dados através de *hashtags*. Apesar desse trabalho não utilizar textos grandes, a metodologia utilizada nos experimentos serviram de inspiração para conduzir a etapa de validação do trabalho proposto por mim.

Seguindo a mesma linha de trabalho de Kouloumpis, Wilson e Moore (2011), houve um outro estudo comparativo para textos do twitter (PSOMAKELIS et al., 2015) que realizou a comparação entre mecanismos para obtenção de análise de sentimento, os quais foram *bag-of-words*, *n-grams* e *n-gram graphs*, e para cada um deles foi realizado uma medição para sete algoritmos de classificação, os quais foram: SVM, *Naïve Bayes*, *Logistic Regression*, *Multilayer Perceptrons*, *Best-First Trees*, *Functional Trees* e *C4.5*. Como resultado, foi mostrado a eficácia da utilização de *n-grams* para o pro-



blema de análise de sentimento em *tweets* e mostrou que é possível realizar a combinação de diferentes métodos a fim de criar novas técnicas para melhoramento dos resultados.

Houve também um estudo comparativo (PADMAJA, 2014) entre a análise de sentimento utilizando o *Naive Bayes*, SVM e um dicionário léxico chamado por *SentiWordNet* (ESULI; SEBASTIANI, 2006) que foi baseado na *WordNet* (MILLER, 1995) e utilizando textos de notícias da Índia mostrou que a análise de sentimento em textos de notícia é difícil por que muitas vezes não é exposto claramente o que o autor do texto quer transmitir e que é preciso um conhecimento sobre o contexto da notícia. Apesar da dificuldade de classificar claramente um texto, foi possível obter a classificar as notícias com uma taxa de 70%. Apesar da taxa de acerto relativamente baixa, o estudo mostrou-se promissor e com uma utilização e teste de outras técnicas, é possível que atinja melhores resultados.

Por último, existe um trabalho para análise de sentimento em textos persas (FARHOODI; YARI; SAYAH, 2011), que trabalhou com a predição de *n-grams*, foi preciso realizar essa etapa de predição porque a linguagem persa é difícil de indentificar palavras. Após a identificação dos *n-grams*, é utilizado classificadores (Naïve Bayes, SVM e KNN) para realizar a predição de um texto para uma determinada classe. Realizando experimentos, o estudo trouxe como resultado uma melhor performance utilizando *3-grams* para textos persas. Como melhoramento, talvez seja possível realizar a classificação de um texto para várias classes ou tópicos. Além disso, a combinação de algoritmos de aprendizado de máquina e *n-grams* pode melhorar no desempenho da classificação textual.

## 2.2 REST API

O nome REST dado por *Representational State Transfer*, originado pela tese de doutorado de Roy Fielding (FIELDING, 2000), cujo trabalho apresentou o modelo de arquitetura REST. O protocolo REST é uma forma de comunicação baseada em cliente-servidor e operações *stateless*. Os serviços cliente-servidor podem gerar uma estrutura de sistemas distribuídos no qual o *server (host)* disponibiliza um ou mais

serviços para que o clientes os acessem. O termo *stateless* dá-se pelo fator de não armazenar dados do cliente, e as informações de sessão são armazenadas do lado do cliente (PAUTASSO; WILDE; ALARCON, 2014).

Para realizar tal tarefa, o cliente envia uma requisição HTTP (GET, POST, PUT, DELETE) para o *endpoint* disponível no *host*. O *endpoint* é o endereço determinado no *server* com uma determinada atividade. Sendo assim, essas atividades podem receber diferentes métodos HTTP, e para fins explicatórios, esse métodos podem ser relacionados com CRUD (*Create Read Update Delete*) utilizado em operações de banco de dados.

Os métodos utilizados pelo REST são classificados como:

- **GET**: Utilizado para consultar um objeto;
- **POST**: Utilizado para inserção ou criação de objetos;
- **PUT**: Utilizado para atualizar um objeto;
- **DELETE**: Utilizado para deletar um objeto.

Como resposta do *server*, é dado como resposta um HTTP *Status Code* padrão. Esses códigos indicam como a requisição foi tratada pelo servidor, e servem para o cliente saber se houve algum problema na requisição informada ou se foi aceita com sucesso. Além do *status code*, o servidor pode realizar operações e enviar como resposta o resultado da operação.

Os *status code* mais comuns em APIs REST são:

- **200**: Requisição aceita;
- **400**: Indica uma *Bad Request*, podendo ser uma falta de parâmetro necessária na requisição;
- **401**: Erro de *Unauthorized*, causado pela falta de autenticação;
- **403**: Erro de *Forbidden*, causado por falta de privilégios do usuário.

- **404:** Erro de *Not Found*, o endereço solicitado não existe.

A escolha de utilizar uma API REST na ferramenta foi devido à facilidade de comunicação, serve como base para criação de outras aplicações ou interfaces futuramente, podendo transformar um projeto inicial em uma plataforma que alimenta diversos projetos.

## 2.3 Mineração de Textos

A Mineração de Textos pode ser relacionada como uma extensão de Mineração de Dados (TAN, 1999). Enquanto que a Mineração de Dados trata de descoberta de conhecimento em dados estruturados, a Mineração de Textos irá seguir as mesmas etapas da Mineração de Dados para descoberta de conhecimento (GOMES, 2013), mas com uma etapa anterior de estruturação dos dados, pois irá tratar com dados não estruturados.

Como o computador apenas interpreta o texto como uma sequência de caracteres, sem possuir uma estruturação, esse volume de dados no formato de texto não pode ser utilizado para extração de conhecimento. Por essa razão são utilizados algoritmos para estruturação do texto de uma maneira que o computador consiga reconhecer. Além de utilizar algoritmos, existem outras maneiras de extração de conhecimento do texto, utilizando outras áreas do conhecimento, como a linguística, em que é possível extrair informações através da formação sintática do texto (WIEBE, 1994).

A Figura 1 mostra um modelo de mineração de texto proposto por Aranha (2007) com cinco etapas distintas: coleta, pré-processamento, indexação, mineração e análise. Na coleta utiliza-se *web crawler*, em que de uma maneira automatizada acessam páginas na *web* através de requisições e extraem informações disponíveis no conteúdo da página. No pré-processamento, é realizada a limpeza do texto, onde são removidas as *stop words*, palavras em que não acrescentam valor ao texto, palavras como: "de", "os", "pelas". Além disso, nessa etapa é realizada a *tokenização* da frase, criando uma lista de palavras em que cada elemento é uma palavra, no caso de *unigram*. Na etapa de indexação, é realizada uma busca por características do texto a fim de obter um

ganho de informação sobre o texto previamente processado, por exemplo a busca de palavras-chave. Na fase de mineração de dados, é determinado qual algoritmo será aplicado ao conjunto de dados com o objetivo de criar um modelo de classificador para realizar a predição da classe correspondente aos dados. Na etapa de análise, os resultados obtidos são avaliados e validados para enfim retirar uma conclusão dos resultados.



Figura 1 – Etapas do processamento de texto  
(ARANHA, 2007)

Este trabalho baseou-se no processo de mineração de dados proposto por [Ara-nha \(2007\)](#), entretanto não será executado o passo de indexação por não haver a necessidade de indexar os textos extraídos, uma vez que já foram previamente indexados durante a etapa de extração.

## 2.4 Processamento de Linguagem Natural

Processamento de Linguagem Natural (PLN) é um conjunto de técnicas computacionais que visa processar dados textuais para diversas tarefas, como identificação semântica e tradução de textos ([CAMBRIA; WHITE, 2014](#)).

Segundo a definição de [Liddy \(2001\)](#), o Processamento de Linguagem Natural é um conjunto de técnicas computacionais motivadas pela teoria de análise e representar naturalmente textos em vários níveis de análise linguística com o propósito de atingir processamento de linguagem próxima à humana em uma série de aplicações ou tarefas. ([LIDDY, 2001](#), p. 2, tradução nossa).

A utilização de PLN envolve elementos léxicos, semânticos e morfológicos da

linguagem (LIDDY, 2001). A complexidade desse processo, se relaciona diretamente com a complexidade da linguagem natural, que pode ser medida pela estrutura gramatical e regras que compõem a linguagem.

No PLN, umas das práticas de obter palavras que acrescentam valor à solução é o pré-processamento, em que o texto passa por uma etapa de limpeza. Essas técnicas de PLN são conhecidas como remoção de *stopwords*, *stemming*, segmentação de palavras, entre outras. *Stopwords* são palavras que não possui um valor para classificação de texto por se repetir bastante. Exemplo de *stopwords* são palavras que servem como conectores, artigos e preposições: e, pois, para. *Stemming* é transformar a palavra em uma variação raiz, em que palavras similares sejam iguais ao *stemming* da palavra em comparação, por exemplo: a palavra quinto possui o *stemming* quint, que seria igual ao da palavra quintos, entretanto diferente da palavra quintal, que seria quintal. Essa técnica de *stemming* é interessante pois possibilita o agrupamento de palavras similares, diminuindo assim a quantidade de palavras utilizadas para classificar o texto.

Nesse trabalho, como os textos das notícias geralmente são grandes, em comparação com textos de *reviews* e *tweets*, a quantidade de palavras obtidas é enorme. Utilizando as técnicas de PLN, como *stemming* e tokenização, é possível diminuir a quantidade de palavras que serão utilizadas e conseqüentemente melhorar o desempenho do classificador.

#### 2.4.1 Análise de sentimento

Trata-se de um campo de pesquisa de mineração de texto em que realiza um estudo computacional sobre opiniões, atitudes e emoções sobre um indivíduo (MEDHAT; HASSAN; KORASHY, 2014). A informação textual pode ser classificada em dois principais tipos: fatos e opiniões. Fatos são expressões objetivas sobre indivíduos, eventos e suas características. Opiniões são geralmente expressões subjetivas que descrevem os sentimentos de uma população sobre determinadas circunstâncias ou indivíduo (INDURKHYA; DAMERAU, 2010). Sendo assim, análise de sentimento é o estudo que determina a polaridade, que podem ser classes como positiva, negativa e neutra, de textos relacionados à algum indivíduo, objeto, local ou circunstância.

Apesar de análise de sentimento ser feita em textos com palavras que trazem sentimento explícito, como excelente, bom e ruim, é possível extrair a polaridade de textos que não demonstram sentimento algum. Pelo fato de que a análise de sentimento ser realizada por classificadores, é possível extrair as palavras de características dos textos e assim inferir a polaridade textual (GOMES, 2013).

## 2.5 Ferramentas

Nesta seção serão apresentadas as ferramentas que foram mais importantes para o trabalho que são: Scrapy, MongoDB e Python NLTK.

### 2.5.1 Scrapy

Scrapy é um *framework* de código aberto para extração de dados em páginas *web*. Escrito em *Python*, e possuindo uma arquitetura baseada em eventos, esse *framework* possui como principal objetivo a extração de dados estruturados a partir de páginas na *web* (KOUZIS-LOUKAS, 2016).

Esse *framework* é utilizado em ferramentas de mineração de dados, processamento de informação e arquivamento histórico. Apesar de ter sido criado para extração de dados, o *Scrapy* pode ser utilizado para outros propósitos, como extrair dados de APIs ou funcionar como um *web crawler* de propósito geral para vasculhar páginas.

#### 2.5.1.1 Arquitetura

A arquitetura do *Scrapy* é baseada em um fluxo de dados e orientada à eventos. O *Scrapy* foi escrito com o *Twisted*, um *framework* em *Python* para redes baseadas em eventos. Sendo assim, é implementado para funcionar de maneira assíncrona, sem bloqueios. Apesar de executar de maneira assíncrona, o *Scrapy* realiza a orquestração de fluxo para manter uma organização dos componentes e prevenir conflitos.

Para realizar o fluxo, o *Scrapy* divide-se em sete componentes, como é possível ver na Figura, são eles (Scrapy developers, 2016):

- **Scrapy Engine:** Esse componente é responsável por controlar o fluxo de dados

entre os componentes do sistema, e ativação de eventos quando determinadas ações ocorrem;

- **Downloader:** É responsável por obter as páginas *web* e enviá-las para a *Engine*, o qual irá encaminhar para as *spiders*;
- **Scheduler:** Recebe as requisições vindas da *Engine* e realiza um enfileiramento para que sejam utilizadas posteriormente quando a *Engine* requisitar;
- **Spiders:** São classes customizadas escritas pelos usuários do *Scrapy* a fim de analisar respostas de requisições e extrair itens delas, ou seguir outras requests em outras *URLs*;
- **Item Pipeline:** Responsável pelo processamento dos itens que foram extraídos pelas *spiders*. Este componente possui tarefas como limpeza, validação e persistência dos itens;
- **Downloader Middlewares:** São vários elementos que se encontram entre a *Engine* e o *Downloader*. Eles funcionam para processar as *requests* e *responses* trocadas pelos dois componentes. Os *Downloaders Middlewares* são úteis quando é preciso realizar alguma modificação na *request* antes de enviar para o *Downloader*, como adicionar um campo na *request* ou *cookies*;
- **Spider Middlewares:** Elementos localizados entre a *Engine* e as *spiders* e são capazes de processar a entrada da *spider* (*responses*) e a saída (*requests*). Esse componente tem utilidade como realizar o tratamento de exceções levantadas pelas *spiders* ou realizar alterações nas *requests* das *spiders*;

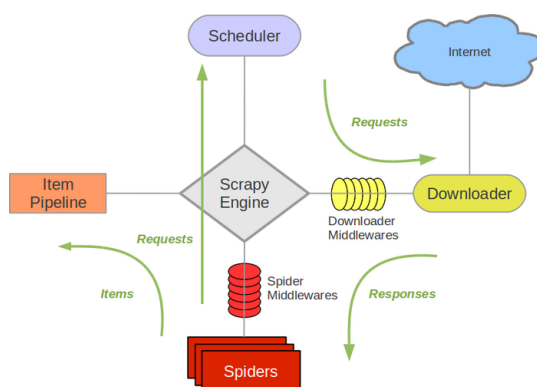


Figura 2 – Arquitetura do Scrapy (Scrapy developers, 2016)

O *Scrapy* possui um fluxo, uma sequência de eventos, e é controlado pela *Engine*, um dos sete componentes da arquitetura do *Scrapy*. O fluxo é descrito da seguinte forma (KOUZIS-LOUKAS, 2016) de acordo com Figura 2:

1. A *Engine* obtêm as primeiras URLs da *Spider* e realiza o agendamento das URLs no *Scheduler*, como *requests*;
2. A *Engine* solicita ao *Scheduler* a próxima URL;
3. O *Scheduler* retorna a próxima URL para a *Engine* e então a *Engine* as envia para o *Downloader*, passando também pelo *Downloader Middleware*, como *request*;
4. Quando o *download* da página finaliza, o *Downloader* gera uma *response* da página e envia de volta para a *Engine*, passando pelo *Downloader Middleware*;
5. A *Engine* recebe a *response* do *Downloader* e a envia para a *spider* realizar o processamento, passando pelo *Spider Middleware* como *input*;
6. A *spider* processa a *response* e retorna os itens obtidos e uma nova *request* para a *Engine*, passando novamente pelo *Spider Middleware* como *output*;
7. A *Engine* realiza o encaminhamento dos itens obtidos para o componente de *Items Pipelines* e a nova *request* para o *Scheduler*;
8. O processo se repete do primeiro passo até que não existam mais *requests* no *Scheduler*;

O *framework Scrapy* foi escolhido para ser utilizado nesse trabalho, pois além de prover várias funcionalidades que irão facilitar bastante a captura de dados para a ferramenta proposta, é um dos principais *frameworks* de extração de dados na *web* na indústria hoje em dia (WANG; GUO, 2012). Além do *scrapy*, existem outros *frameworks* para *scraping*, como o *BeautifulSoup*, *lxml* e até mesmo o *requests* se utilizado em conjunto com alguma outra biblioteca para extrair *XPaths* de uma página.



## 2.5.2 MongoDB

Trata-se de um banco de dados open-source e que possui um crescimento e aceitação bastante alto (STONEBRAKER, 2010). Além disso, o MongoDB possui características como flexibilidade no modelo de dados, onde o *schema* dos dados podem ser alterados facilmente, por não possuir validadores de dados sofisticados nos documentos. Por possuir esse modelo de dados flexível, o MongoDB, consegue lidar com as evoluções da aplicação, como mudança e inserção de novos modelos, ao contrário de bancos relacionais que precisariam de uma remontagem da tabela no banco, podendo causar erros (YOONA et al., 2016).

Por ser NoSQL, a estrutura do MongoDB é orientado à documentos e diferentemente de bancos relacionais, como o Postgres, ele não possui relacionamento entre os documentos e não existe o conceito de *Constraint*. Apesar disso, bancos NoSQL foram criados para serem utilizados com aplicações de *Big Data*, pois banco de dados relacionais não possuem um desempenho tão bom quando é envolvido uma grande quantidade de dados (CHICKERUR; GOUDAR; KINNERKAR, 2015).

Os documentos salvos no MongoDB são no formato binário BSON (*Binary JSON*); Esse formato é uma extensão do formato JSON com o acréscimo de outros tipos de dados, como *int*, *long*, *date* e *floating point*. No formato BSON, o documento pode possuir um ou mais campo e cada campo pode conter um específico tipo de dado, *array* ou até outro documento.

Para realizar consultas, o MongoDB oferece outras formas de busca fora a *Key-value*, onde é filtrado o resultado utilizando um campo e um valor. Existem outras formas de realizar consultas mais complexas, como por exemplo:

- **Busca em texto:** Utilização de expressões regulares para realizar uma busca no texto de algum campo específico;
- **MapReduce:** É realizado um processamento de dados através de uma expressão definida em *Javascript* e executada nos dados armazenados no banco;
- **Aggregation framework:** Agregação de valores que foram encontrados através

de outras funções, como o *count*, *sum*, *min* e *max*. Além disso, utilizando o *\$lookup*, é possível realizar a combinação de documentos que se encontram em diferentes coleções.

A escolha desse modelo de banco de dados para a ferramenta proposta foi devido à disponibilidade de ferramentas que são executadas a nível de banco de dados que irão auxiliar na implementação. Além disso, será preciso realizar consultas no texto e o MongoDB oferece um ótimo desempenho comparado com outros bancos de dados convencionais (CHICKERUR; GOUDAR; KINNERKAR, 2015). Nesse estudo, foi mostrado um comparativo entre o MongoDB e MySQL para operações de *insert*, *update*, *select* e *delete*, no qual, o MongoDB executou em menos tempo as operações citadas, principalmente envolvendo grande volume de dados.

### 2.5.3 Python NLTK

O Python Natural Language Toolkit é uma biblioteca de código aberto organizado em módulos. Essa biblioteca foi criada para fins pedagógicos com o intuito de tornar a criação de novos projetos de maneira mais prática, sem que haja a necessidade de implementar funcionalidades básicas e que o foco seja no assunto (LOPER, 2004b). As funcionalidades básicas que essa ferramenta provê são bibliotecas de processamento de texto, como stemming, parsing, classificação, tokenização, tratamento semântico e tagueamento, oferecendo suporte não apenas para o inglês, mas também para o português.

Essa ferramenta é utilizada em aplicações de tratamento de texto, como por exemplo, análise de sentimento em discursos e mineração de opiniões do público em publicações (KOTHAPALLI; SHARIFAHMADIAN; SHIH, 2016). Além disso, por ser uma ferramenta de código aberto e possuir uma comunidade bastante presente, fez com que o Python NLTK se tornasse uma das principais ferramentas de tratamento de texto em Python.

## 3 Metodologia

Este capítulo apresenta a metodologia e o método de pesquisa adotado neste trabalho. A metodologia foi baseada no fluxo proposto por (ARANHA, 2007), inicialmente será feita a coleta dos textos, e então os textos obtidos passarão por um pré-processamento para que sejam utilizados pelo classificador de textos que irá ser treinado para que seja possível realizar a análise sentimental pela predição da classe em que o texto pertence.

### 3.1 Método de pesquisa

A [Figura 3](#) mostra as etapas de pesquisa utilizadas no trabalho, onde inicialmente foi realizado uma revisão da literatura com o objetivo de encontrar trabalhos relacionados e maneiras de trabalhar com textos para classificação sentimental. Em seguida, foi escolhido um modelo de etapas para mineração de texto, que foi selecionado a partir do critério de similariedade com o trabalho proposto. Após escolher o modelo, teve-se início o desenvolvimento da ferramenta, que gerou documentos como a arquitetura e requisitos do *software*, além da codificação inicial da ferramenta.

Após o término do desenvolvimento da ferramenta, houve a realização de experimentos para determinar a % da base de dados que será utilizada para treinamento e teste. Além disso, nessa etapa foi realizado também a determinação do intervalo de palavras que farão parte do conjunto de palavras de característica. Por fim, foi realizado uma análise e avaliação da ferramenta, produzindo gráficos e resultados que são discutidos no Capítulo 5 deste documento.

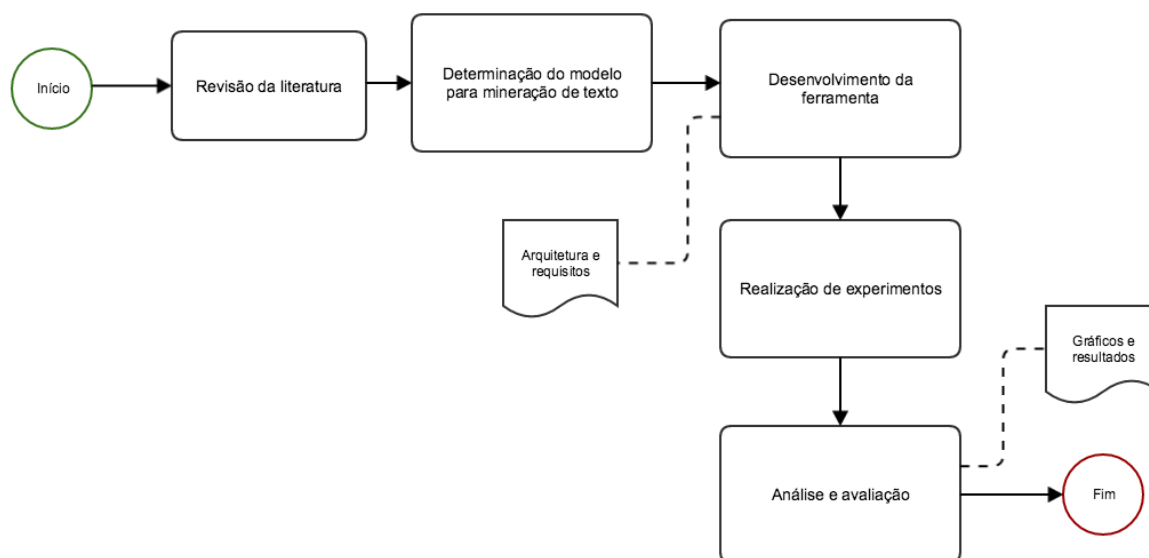


Figura 3 – Método de pesquisa utilizado no trabalho

O método científico utilizado nesse trabalho é descrito de natureza aplicada, com objetivos explicativos, abordagem qualitativa e procedimento de estudo de caso, devido à geração de produto de maneira imediata utilizando dados de um determinado grupo ou indivíduos, que serão analisados, classificados e interpretados para retirada de conclusões.

### 3.2 Determinação do modelo para mineração de texto

Nesta seção será mostrado como foi realizado cada etapa do modelo de mineração de texto mostrado na [Figura 4](#) inspirado por [Aranha \(2007\)](#).



Figura 4 – Modelo de mineração de dados inspirado em [Aranha \(2007\)](#)

### 3.2.1 Coleta de texto de notícias

A coleta dos textos das notícias foi a primeira etapa de execução desse trabalho. Nessa etapa foi utilizada a criação de *web crawlers* (*spiders*) utilizando a linguagem Python junto com o *framework* Scrapy. Foram criadas *spiders* para um número mínimo para validação da proposta do trabalho. A *spider* recebe como parâmetro o nome do indivíduo que deseja ser consultado e retorna as notícias relacionadas à pessoa de acordo com a *spider* solicitada.

Apesar de se tratar de uma etapa de coleta, os textos passaram por uma etapa de processamento, em que foram removidos os *hyperlinks* que são encontrados dentro do conteúdo do texto. Esse processamento foi necessário para evitar *tags* do HTML da página dentro do conteúdo que será utilizado nas próximas etapas. Além disso, foi realizado um padronização na data de publicação capturada para evitar diferentes formatações de datas, como por extenso e numeral.

Por se tratar de uma escala nacional, em que os canais de notícias escolhidos relatam notícias do Brasil inteiro, os indivíduos escolhidos como objetos de pesquisa são participantes do Governo Federal e possuem cargos de Presidente, Senador e Deputado Federal. Para a busca destes indivíduos, os canais de notícias escolhidos para estudo foram G1, O Globo, Época, Carta Capital, UOL e Folha de São Paulo. A escolha dessas mídias foi seguindo o critério de importância, credibilidade, popularidade e disponibilidade para *scraping* utilizando o *framework* Scrapy. Apesar de existirem outras mídias importantes, foi selecionado apenas um escopo reduzido para estudo.

A [Figura 5](#) apresenta um exemplo de notícia salva no banco de dados (MongoDB) da ferramenta. Nesse trabalho são extraídos dados que foram considerados relevantes para o propósito da ferramenta, como URL, data de publicação, título, conteúdo e autor. Outros dados foram salvos para controle, como data de extração e fonte da notícia.

_id	ObjectId("58881c9e7f6afa450d6e7d76")	ObjectId
titulo	Lupi vai ao Senado prestar esclarecimentos sobre uso de jatinho	String
extraido_em	2017-01-25 12:46:40.014Z	Date
data_publicacao	17/11/2011	String
url	http://www.cartacapital.com.br/politica/lupi-vai-ao-senado-prest...	String
autor	Agência Brasil	String
conteudo	Alvo de denúncias desde a semana passada, o ministro do Trabal...	String
fonte	cartacapital	String
indivduo	[ 1 element ]	Array
[0]	Dilma Rousseff	String

Figura 5 – Exemplo de notícia armazenada no banco de dados

Após a finalização desta etapa, iniciou-se a etapa de pré-processamento, que é explicada na subseção a seguir.

### 3.2.2 Pré-processamento

Utilizando os textos das notícias obtidos na coleta, foi aplicado as técnicas de PLN, os quais foram: tokenização, remoção das *stopwords* e *stemming*. Para realizar a tokenização e *stemming*, foi utilizado a biblioteca Python NLTK, quanto às *stopwords*, foi utilizado uma extensão das *stopwords* do Python NLTK, em que foi adicionado novas palavras que não estavam na lista de *stopwords* da biblioteca. O objetivo desse pré-processamento é realizar a limpeza das palavras que possuem pouca importância para a classificação do texto e irá facilitar a obtenção das palavras de característica do texto.

Todos os textos que foram coletados irão passar pela etapa de pré-processamento. A diferença é que para o conjunto de treino e teste serão criadas listas de tuplas<sup>1</sup>, que serão as palavras junto com a classificação do texto. Essa lista de tuplas será posteriormente utilizada para realizar o treinamento e teste do classificador.

Os textos da base que não pertencem ao conjunto de treino e teste precisam passar pela etapa de pré-processamento para realizar a padronização das palavras. Essa padronização é importante para a predição do texto.

<sup>1</sup> Tuplas consistem em um número de valores separados por vírgula e entre parênteses, além disso, possuem a propriedade de serem imutáveis. Ex.: ('exemplo', 2) (FOUNDATION, 2017b)

### 3.2.3 Classificação dos textos

Nessa etapa foram utilizados classificadores da biblioteca *scikit learn*. Essa biblioteca é utilizada para aprendizado de máquina em projetos acadêmicos ou até mesmo em projetos comerciais (LOPER, 2004a). Foi utilizado três classificadores diferentes para esta ferramenta, *Naive Bayes*, *SVM Linear* e *SVM* utilizando *kernel* de função radial (RBF) (SCHOLKOPF et al., 1997). A escolha desses classificadores foram baseadas em trabalhos anteriores de análise de sentimento com a mesma problemática do trabalho proposto. Esses classificadores possuíram resultados satisfatórios (PADMAJA, 2014) em análise de sentimento de textos de notícias. Com relação aos classificadores, não houveram alterações nos parâmetros iniciais para inicialização dos classificadores utilizados a etapa de treinamento, pois não foram realizados testes de performance .

Utilizando um conjunto de dados limitado de notícias, foram separadas as notícias para o conjunto de treinamento e teste, em que 5% do conjunto total de notícias foi reservado para o conjunto de teste, e 30% do conjunto de treinamento foi utilizado para realizar teste, o restante foi utilizado para treinar o classificador. As notícias foram classificadas manualmente em três categorias, neutra, positiva e negativa. As notícias foram classificadas seguindo alguns critérios, como descrito na Tabela 1.

	Critérios
Positivas	Criação de leis que ajudam a população, inauguração de obras, participação em novo cargo, liderança nas pesquisas eleitorais.
Negativas	Participação em escândalos, lava-jato, compra utilizando dinheiro público sem licitação e perda de popularidade.
Neutras	Notícias de pouca importância, como por exemplo, mudança de residência. Além disso, notícias em que o indivíduo é apenas citado de maneira que a notícia não é relacionada diretamente com ele.

Tabela 1 – Critérios utilizados para classificação de notícias.

Houveram casos em que foram extraídas notícias dos indivíduos buscados, mas os mesmos não são citados no texto. Esse problema foi notado com as notícias referentes à Carta Capital. O motivo deste problema é desconhecido e existe uma hipótese de que tenha se originado pelo critério de busca interno da página da Carta Capital.

Após a finalização desse processo, teve início o processo de avaliação dos resultados obtidos dos classificadores, que será descrito com mais detalhes na próxima sessão.

### 3.2.4 Análise de sentimento

Com o objetivo de obter as palavras de características, foram utilizadas metodologias de PLN para montar os dados que seriam recebidos pelo classificador. Após a etapa de pré-processamento, foi gerado um conjunto de todas as palavras contidas na base de treinamento. utilizando esse conjunto de palavras, foram filtrados apenas as palavras que tinha ocorrência menor que 55 e maior que 9 em todos os textos da base de treinamento, essa escolha foi realizada de maneira arbitrária com o objetivo de obter palavras que repetissem em todos os documentos de forma mediana, em comparação com as palavras de maior repetição e a quantidade de documentos. Ao obter essas palavras de características, foi criada uma matriz em que cada linha é uma notícia e cada coluna é uma tupla com uma lista de palavras do conteúdo da notícia, seguido de um binário que determina se a palavra está no conjunto de palavras de características. No final de cada linha, existe uma coluna que indica qual a polaridade da notícia, que foi pré-determinada anteriormente. A Figura 6 mostra o conjunto de palavras de características, onde o primeiro elemento da tupla é a palavra e o segundo é a quantidade que se repete em todos os textos.

```
[('políticos', 55), ('poder', 55), ('mulher', 54), ('então', 54), ('pesquisa', 54), ('tempo', 53), ('o', 52), ('michel', 52), ('dinheiro', 52), ('votação', 52), ('político', 51), ('menos', 50), ('pl', 'parte', 49), ('outros', 48), ('fim', 48), ('base', 48), ('direito', 48), ('programa', 48), ('ant', 'missão', 47), ('aécio', 47), ('contas', 47), ('onde', 47), ('eleições', 46), ('além', 46), ('rela', 'D', 'pedido', 46), ('discurso', 45), ('sido', 45), ('psdb', 45), ('palácio', 45), ('justiça', 44), 'a', 44), ('afastada', 44), ('operação', 43), ('julgamento', 43), ('primeira', 43), ('semana', 42), 'a-feira', 41), ('todo', 41), ('mil', 41), ('voto', 41), ('tudo', 40), ('sendo', 40), ('crise', 40), '39), ('brasileiros', 39), ('ação', 39), ('cada', 39), ('novo', 39), ('eleição', 39), ('cunha', 'selho', 38), ('três', 38), ('união', 38), ('feito', 37), ('civil', 37), ('desta', 37), ('candidat', 'D', 'deputado', 37), ('banco', 36), ('entrevista', 36), ('acordo', 35), ('economia', 35), ('sempre', 'uas', 35), ('nota', 35), ('sociais', 35), ('deste', 35), ('segunda-feira', 35), ('qualquer', 34), '34), ('rio', 34), ('ato', 33), ('neste', 33), ('brasileiro', 33), ('forma', 33), ('vida', 33), ('', 'bom', 33), ('atual', 32), ('joão', 32), ('advogado', 32), ('início', 32), ('afirma', 32), ('p', 'D', 'favor', 32), ('2', 32), ('16', 31), ('afastamento', 31), ('popularidade', 31), ('poderia', 31), ('3', 'mulheres', 31), ('parecer', 31), ('nunca', 31), ('especial', 31), ('fernando', 31), ('sociedade
```

Figura 6 – Exemplo de conjunto de palavras de características

Para técnicas que utilizam aprendizado de máquina realizar análise de senti-



mento, a forma de treinamento utilizado é geralmente o supervisionado (VINODHINI; CHANDRASEKARAN, 2012), em que é passado um conjunto de dados previamente classificados e o algoritmo realiza ajustes dos pesos para atingirem o objetivo de classificação, caso a resposta informada pelo classificador seja diferente da resposta no exemplo, é realizada uma correção nas constantes de pesos do classificador.

Utilizando os classificadores que geraram os modelos através da base de treinamento, é possível determinar a polaridade de uma notícia arbitrária. A fim de determinar a taxa de acerto dos classificadores, foi utilizada a base teste e então foi comparado o resultado informado pelo classificador e a resposta que foi previamente informada manualmente.

Como cada classificador pode informar uma resposta diferente, o resultado final da classificação do texto é dada pela classe mais presente nas respostas, por exemplo: dois classificadores informaram que o texto pertence à classe positiva e um para a negativa, a resposta será a classe positiva. Entretanto, por possuir três classes e três classificadores, pode haver o caso de ser informado uma resposta diferente para cada um. Nesse caso, a resposta final será do classificador com maior taxa de acerto., por exemplo, o texto é classificado como Negativo, Positivo e Neutro por cada um dos classificadores, a resposta final será dada pelo classificador com maior taxa de acerto.

A análise de sentimento das notícias é um processo complexo em que é preciso analisar ponto-de-vista diferentes, em que a mesma notícia pode ter efeitos positivos para determinados indivíduos e negativo para outros. Por causa dessa complexidade, tornou-se necessário o treinamento diferente para cada indivíduo, pois existem notícias em que se encaixam no quadro descrito anteriormente.

Por fim, os passos descritos nesta seção foram importantes para o desenvolvimento da ferramenta, pois indicam as etapas que o trabalho proposto deve seguir e as técnicas que serão utilizadas nos próximos capítulos.

## 4 Desenvolvimento da Ferramenta

Neste capítulo serão apresentados as seções de levantamento de requisitos, a maneira de como a ferramenta foi arquitetada e os testes realizados.

### 4.1 Levantamento de Requisitos

Os requisitos para um projeto de *software* são importantes para manter objetivos durante a etapa de desenvolvimento (POHL, 2010). Requisitos definem os serviços que a ferramenta irá oferecer e as restrições com relação às operações. Segundo Kandt (2003), os requisitos podem ser classificados em:

- Requisitos funcionais (RF): declarações de funcionalidades que o sistema deve oferecer. As funcionalidades estão relacionadas com operações realizadas pelo usuário ou o próprio sistema, execuções de tarefas internas e externas à ferramenta.
- Requisitos não funcionais (RNF): indicam restrições nas operações do sistema. Essas restrições se referem a limitações de requisições, limites de tempo de resposta, desempenho, custo.

Apesar de que os requisitos estão em categorias diferentes, eles podem exercer influência nos outros, por exemplo, um requisito funcional quando afeta o desempenho da ferramenta, pode prejudicar a disponibilidade do serviço, caso este esteja classificado como um requisito não-funcional.

Os requisitos funcionais foram classificados em Captura, Mineração de texto e API. A classificação de Captura foi devido ao relacionamento com a obtenção dos textos e a busca por indivíduos no canal de notícia. Requisitos funcionais relacionados ao tratamento dos textos obtidos e treinamento dos classificadores são classificados como Mineração de texto. Por fim, requisitos funcionais que apresentam comportamento relacionado à interface entre o usuário e a ferramenta, foram classificados como API.

Já em relação à prioridade, foram utilizados prioridades alta, média e baixa, requisitos com prioridade alta são aqueles que a ferramenta não conseguiria funcionar corretamente sem eles, prioridade média são requisitos que irão ajudar no funcionamento, mas não são exatamente necessários para o funcionamento básico e prioridade baixa são os requisitos que irão acrescentar algo à ferramenta, mas o ganho de funcionalidade é baixo, comparado com os requisitos de prioridade alta e média.

O detalhamento dos requisitos, como descrição e prioridade, encontram-se no Anexo A.

#### 4.1.1 Requisitos Funcionais

Os requisitos funcionais foram separados em duas categorias. A categoria de captura trata-se dos requisitos das *spiders*, que irão acessar as páginas de busca dos jornais e capturar os dados da notícia. A categoria de *core* trata-se da aplicação que irá processar os dados capturados durante a etapa de coleta de dados.

- Captura:
  - RF01: Buscar por indivíduo;
  - RF02: Obter texto de notícia;
  - RF03: Tratamento de texto e padronização;
  - RF04: Persistência de notícia;
  - RF05: Armazenar dados da captura.
- Mineração de texto:
  - RF01: Realizar pré-processamento de texto;
  - RF02: Treino do classificador por indivíduo;
  - RF03: Predição da classificação do texto.
- API:
  - RF01: Método GET para obter informações sobre um indivíduo;
  - RF02: Método POST para realizar a procura de indivíduo;

### 4.1.2 Requisitos Não Funcionais

- RNF01: Segurança dos dados;
- RNF02: Desempenho;
- RNF03: Disponibilidade do serviço.

## 4.2 Arquitetura

O projeto foi dividido em diferentes componentes que são: captura, classificação e *view*. A captura é onde as *spiders* estarão disponíveis e podem ser agendadas para execução através de um controlador de tarefas próprio (*Scrapyd*) via API. O componente de classificação é onde a lógica de negócio do projeto está localizada, nesse componente é realizado o treinamento inicial dos classificadores, a etapa de pré-processamento e análise de sentimento. O componente de *view* funciona como porta de entrada para os usuários, nele localiza-se a implementação da API em REST e comunica-se diretamente com o componente de classificação para realizar o tratamento do texto e retornar uma resposta para o usuário solicitante.

A Figura 7 ilustra como a arquitetura do projeto está organizada. O modelo utilizado foi baseado no padrão MVC (*Model View Controller*) com algumas modificações. O componente responsável pela captura possui uma modelagem inicial dos dados que serão inseridos e com tais dados coletados é realizado um processamento inicial para remoção de texto que não faz parte do conteúdo da notícia e padronização de campos, por exemplo a data de publicação. O componente de classificação, por ser uma unidade que irá orquestrar o fluxo da aplicação e realizará operações com os classificadores, é inspirado no *Controller* do padrão MVC, por ser uma unidade centralizada e possuir a lógica de negócio da aplicação. Por fim, o componente de *View*, é a comunicação com o usuário e por onde as demandas são recebidas, através de uma API REST utilizando o *tornado* e *asyncio*, ambos escritos em Python 3, onde o *tornado* possui suporte à criação de APIs e o *asyncio* realiza atividades assíncronas que serão executadas com a chamada da API.

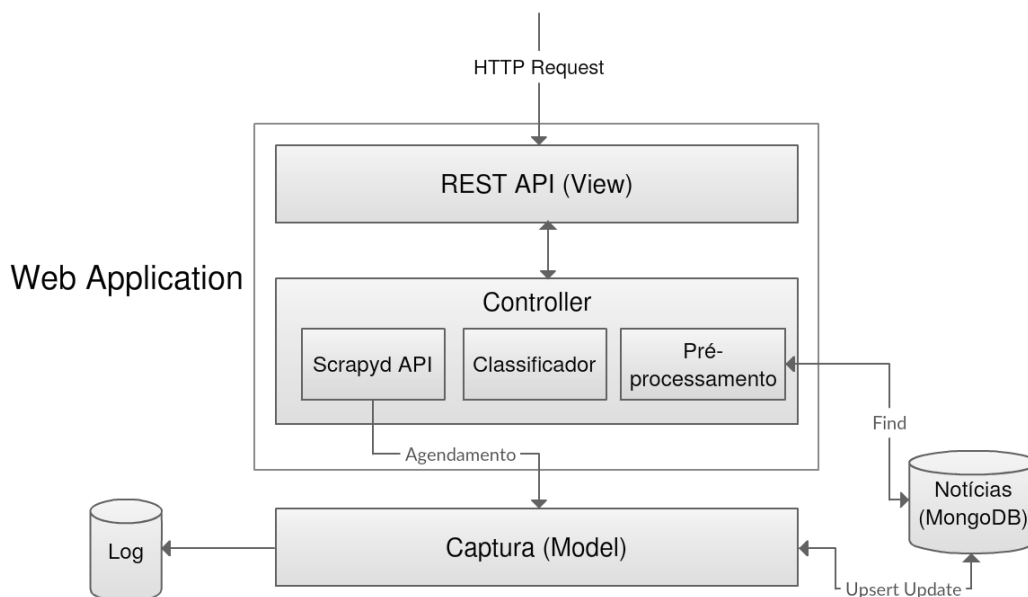


Figura 7 – Arquitetura da Ferramenta Proposta

O *Controller* realiza consultas à aplicação de captura através de uma API. Entretanto, essa API é criada por um serviço de controle do agendamento e execução das *spiders*, possibilitando a operação de múltiplas *spiders* simultaneamente. Esse serviço é chamado de Scrapy, em que a funcionalidade é armazenar os logs das *spiders* e agendar as execuções.

Para obtenção de dados, serão utilizados *web crawlers* para coletar informações dos artigos, visto que alguns portais de notícias não possuem *webservices* de graça. Além disso, estudos realizados com *crawlers* inteligentes com aprendizado de máquina e integração com mineração de dados (DARSHAKAR, 2015), mostraram uma nova maneira de utilizar o *crawler* mais eficiente e foi observado que é possível melhorar a performance e obter resultados mais inteligente. Sendo assim, as informações recebidas após a mineração estariam mais precisas e com uma busca mais eficiente.

Em vista de otimizar a busca, existem técnicas para *web crawler*, como o *MapReduce*, que auxiliam na velocidade de obtenção de dados na web em diminuir o escopo de busca para dados de interesse, retornando assim, apenas os padrões de dados associados à busca do usuário (LEUNG; MACKINNON; JIANG, 2014).

A ferramenta proposta foi construída utilizando Python 3.4 para a API e Contro-

lador. Foi necessário utilizar esta versão do Python para o uso do *asyncio*<sup>1</sup> e utilização de bibliotecas como *scikit-learn* e *numpy*, que são requisitos do Python NLTK.

Para a captura, foi utilizado Python 2.7, pois o serviço de agendamento de *spiders*, *scrapy*, não possui suporte para Python 3.\*. Na etapa de pré-processamento, o Python NLTK foi essencial para tratamento das palavras e tokenização e *stemming*.

Enquanto que para a classificação, foram utilizados classificadores do *scikit-learn*, uma biblioteca em Python bastante utilizadas para ferramentas de aprendizado de máquina. Como comunicador entre as *spiders* e o Controlador, foi utilizado a API do *Scrapy*, em que permite o agendamento de execuções das *spiders*. Apesar do projeto funcionar em versões de Python diferente, não houve conflito de versão ou de código, pois cada módulo do projeto, *spiders*, controlador e *view* ficaram separados em *containers*, utilizando o *Docker* (INC., 2015), onde cada um possui uma estrutura e requisitos próprios, sem conflitar com outras aplicações localizadas em outros *containers*.

### 4.3 Testes Unitários

Os testes foram realizados utilizando um conjunto de notícias previamente classificados manualmente. Os textos das notícias seguiram o fluxo de pré processamento, entretanto, diferentemente do conjunto de treino, as palavras não formaram tuplas com a classe (neutra, positiva e negativa). Para obter a taxa de acerto, para cada classificador, foi realizada a predição dos textos da base de teste e então foi verificado se a classificação foi igual à classe indicada manualmente. O resultado final dos testes dos classificadores mostrou uma taxa de acerto para o Naive Bayes de 51%, Linear SVM 57% e RBF SVM com 55%, esses resultados foram aproximados para melhor entendimento.

Com isso, nesse capítulo foi apresentado os requisitos, a arquitetura da ferramenta e os testes realizados nos classificadores. Além disso, cada componente da ferramenta foi descrito informando a tecnologia utilizada e o objetivo de cada um. No próximo capítulo serão mostrados os resultados obtidos com os indivíduos separados para estudo.

---

<sup>1</sup> Módulo utilizado pelo Python 3.\* para execução de tarefas assíncronas (FOUNDATION, 2017a)

## 5 Resultados

Este capítulo descreve a análise dos resultados obtidos. Os dados utilizados neste capítulo foram capturados seguindo as etapas descritas no capítulo três de metodologia.

### 5.1 Dados coletados

Foram coletados notícias de alguns dos principais canais de notícias do Brasil (G1, O Globo, Época, Carta Capital, UOL e Folha de São Paulo). A escolha desses canais de notícia foi realizada pela relevância e impacto da mídia e a capacidade de realizar uma busca automatizada na *web*. Para o estudo, foram selecionados dois políticos brasileiros, Dilma Rousseff e Renan Calheiros. A escolha desses políticos foi devido à grande polêmica que os envolve no cenário político brasileiro.

	Carta Capital	UOL	Folha de SP	G1	O Globo	Época
Dilma Rousseff	4059	96	122	29	35	13
Renan Calheiros	380	78	119	46	17	55

Tabela 2 – Captura total de notícias

A [Tabela 2](#) mostra a quantidade de notícias capturadas por canal de notícia. Pelo fato de possuir uma grande diferença entre as notícias capturadas pela Carta Capital e os demais canais, foram utilizadas apenas notícias durante o período de Janeiro de 2016 até Janeiro de 2017. Apesar desta restrição, alguns canais de notícias, como Época, O Globo e G1, só foi possível extrair notícias no período de Outubro de 2016 até Janeiro de 2017.

Apesar da filtragem, infelizmente algumas notícias não foram possíveis de serem coletadas utilizando a ferramenta automatizada, pois o conteúdo da notícia de algumas páginas são criados de maneira dinâmica, via *javascript*, por exemplo. Sabendo disso, existe uma limitação na ferramenta de captura utilizada, em que não consegue executar código *javascript*, impossibilitando a aparência do conteúdo para captura. Este artifício é conhecido para possuir uma segurança de conteúdo nas pá-

ginas (PATIL; FREDERIK, 2016) e um exemplo de canal de notícia que utiliza esse artifício é a Veja.

Outros motivos para a baixa quantidade de notícias capturadas no G1, O Globo e Época, por exemplo, foi a grande variação de domínios de acesso, a aparição notícias não relevantes na busca e limitação na quantidade de notícias buscadas. A variação de domínio significa uma quantidade de diferentes publicadores de notícia. O problema principal é a identificação de tais domínios e realizar a captura dos dados através dos XPath. Além disso, algumas páginas não são relevantes para o estudo, visto que podem ser páginas de vídeo ou pertencem à um artigo de opinião ou blog. Quanto à quantidade de notícias disponíveis para captura, existem casos de limitações na consulta, em que são mostrados apenas notícias de um certo período de tempo e problema na paginação, como aconteceu com o UOL.

Com o escopo de notícias estabelecido, os textos coletados pelas *spiders* passaram por uma etapa de processamento básico, utilizando a *pipeline* da arquitetura do Scrapy, onde foram removidos as tags em HTML e os *hyperlinks* inclusos no texto da notícia, pois não atribuem informação para a ferramenta. Outro processamento foi a padronização das datas no formato DD/MM/YYYY HH:MM. Tal padronização da data é importante para a realização de consultas no banco e conseguir visualizar um histórico de notícias.

## 5.2 Pré-processamento

Após a coleta dos dados, foi realizada a leitura de 10% das notícias com mais de 400 palavras para construir uma base de treinamento. A escolha das notícias com uma quantidade maior que 400 palavras teve como motivo buscar notícias que possuíssem um volume grande de texto para que pudesse extrair as palavras de característica de maneira mais abrangente.

A construção da base treinamente foi realizada com a leitura e classificação manual e as notícias foram classificadas em positiva, negativa e neutra em relação ao indivíduo que foi buscado. Esses dados serão úteis para a etapa de classificação, pois o treinamento e teste do classificador será utilizando tais notícias.



Além da leitura, a etapa de pré-processamento realizou a tokenização, remoção das *stopwords* e stemming das palavras restantes. O resultado disso foi uma matriz com cada palavra e a respectiva classe que a identifica.

### 5.3 Validação do modelo de classificação

Após a etapa de pré-processamento, foram coletados 200 notícias seguindo um critério de possuir pelo menos mais do que 150 palavras. Esse critério foi necessário para que as notícias que iriam servir para a base de treinamento sejam grandes e que abranjam uma grande quantidade de palavras, para que possa ser extraído as palavras de característica de cada classe (positiva, negativa e neutra). As notícias desse conjunto separado foram classificadas manualmente pelo autor em uma das três classes criadas. Em seguida, esse conjunto de treino foi dividido em um conjunto de teste (20% da base separada) e um conjunto de treinamento (80% da base separada). Sendo assim, o conjunto de treinamento separado foi utilizado para treinar os três classificadores escolhidos. As notícias selecionadas para treinamento estão disponíveis no Anexo B.

Para escolha das palavras de características, foram criados duas listas. A primeira lista possuem todas as palavras de todas as notícias. Essas palavras passaram pela etapa de pré-processamento, removendo as *stopwords* e normalizando para *lowercase*. Além disso, foram sendo adicionados em uma outra lista, o texto inteiro da notícia em *lowercase* junto com a categoria atribuída ao texto. Possuindo essas listas, para cada notícia, foi verificado a existência das palavras e cada palavra ficou associada à uma classe, de acordo como fosse sendo encontrada no texto. Sendo assim, foi criado uma lista de tuplas com as palavras e um valor binário para determinar se pertence à classe da notícia, como mostra na [Figura 8](#).

```
'argumentos': False, 'construída': False, 'nêmesis': False, 'claque': False, 'escreveram': False, "sacra
mentado": False, 'gordo': False, 'ocupação': False, 'cerimônia': True, 'violados': False, 'observação':
False, 'hostil': False, 'escrito': False, 'ampliarão': False, 'expediente': False, 'contra': False, 'abra
mo': False, 'governador': True, 'presenças': False, 'outubro.': False, 'processuais': False, 'oportuno':
False, 'nenhuma': False, 'epistemológico': False, 'aumenta': False, 'país': False, 'fraga': False, 'boqu
inhas': False, 'gerou': False, 'empreiteira': False, 'finais': False, 'citação': False, 'uruguai': False,
'recorde': False, 'mensagens': False, 'redução': False, 'agradecendo': False, 'acadêmico': False, 'compl
ementares': False, 'irritou': False, 'recebi.': False, 'disposição': False, 'gourmet': False, 'indicação
': False, 'kappaz': False, 'genocídio': False, 'regiões': False, 'lembrado': False, 'juízes': False, 'inf
ormar': False, 'deprimido': False, 'intimidados.isso': False, 'extremamente': False, 'rapidez': False, 'm
ostram': False, 'definido': False, 'danos': False, 'desejou': False, '2010. "está': False, 'afaste': False
, 'convidadas': False, 'fronteira': False, 'reproduzia': False, 'rousseff': False, 'rosa': False, 'edgar
d': False, 'donizeti': False, 'portadores': False, 'costa': False, 'irretocável': False, '1970': False, '
contam': False, 'concessões': False, '2014. "utilizar': False, 'jurídicas': False, 'acenda': False, 'enten
didos': False, 'reincidência': False, 'consolidar': False, 'menino': False, 'gilberto': False, 'rossetto'
: False, 'talita': False, 'influentes': False, 'acertar': False, 'voltam': False, 'intrínseca': False, 'i
ncisivo': False, 'tributaria': False, 'stuckert': False, 'catraca': False, 'livre.e': False, 'relacionad
os': False, 'pires': False, 'ganharia': False, 'tomaremos': False, 'absolver': False, 'construir': False
, 'mencionadas': False, 'dupla': False, 'passarão': False, 'populi/band/ig': False, 'sinistra': False, '
cobrados': False, '43,7': False, 'brutalidade': False, 'apura': False, 'cabos': False, 'rodoviário': Fals
e, 'milhares': False, 'disputado': False, 'fácil': False, 'indignado': False, 'postos': False, "ficha":
False, 'vetar': False, 'data': False, '9h15': False, 'décadas': False, 'greenpeace': False, 'aterrorizant
es': False, 'esforço': False, '19h30': False, 'convidou': False, 'cliente': False, 'inquérito': False, 'd
uque': False, 'totalizando': False, 'propor': False, 'prestigiado': False, 'apagão': False, 'lulopetismo'
: False, 'salvaguarda': False, 'truculência': False, 'venezuelana': False, 'envolvidos': False, 'sd-sp':
False, 'queria': False, 'fugidio': False, 'bruscamente': False, 'diz.astrini': False, 'precisaram': False
, 'querem': False, 'torcem': False, 'pos')
```

Figura 8 – Exemplo de palavras de características de uma notícia

Após os modelos dos classificadores serem gerados, foi verificado a taxa de acerto de cada classificador com o conjunto de teste. O modelo apresentou a taxa de acerto como mostrada na Tabela 3. O classificador SVM Linear apresentou bons resultados com o quadro geral como mostra na Figura 9, com 57% de acerto, junto com *kernel* de RBF, ficou com 55%. Entretanto, apesar do classificador Naïve Bayes ficar com um resultado abaixo em relação com os outros dois classificadores, com 51%, ele foi utilizado para a classificação das notícias. Esses resultados mostram a taxa de acerto para todas as classes, não limitando apenas para positivas, negativas ou neutras.

```
In [3]: from webapp.tasks import nlp_train_data
In [4]: nlp_train_data()
Bayes acc: 0.5147058823529411
SVM Linear acc: 0.5735294117647058
RBF SVC acc: 0.5588235294117647
```

Figura 9 – Taxa de acerto obtida através da base de teste

	Naïve Bayes	SVM Linear	RBF SVM
Taxa de acerto	51%	57%	55%

Tabela 3 – Taxa de acerto dos classificadores

Para teste dos classificadores, foram utilizados 179 notícias, sendo 65 positivas, 99 negativas e 16 neutras. Do total coletado para treino, 200 notícias, 21 foram descartadas por não ser possível determinar a polaridade ou por possuir texto incompleto, ocasionado por erro na captura.

Durante a etapa de classificação manual, foi observado a utilização da omissão de palavras com sentimento ruim ou bom para determinado indivíduo. Além disso, uma mesma notícia pode possuir mais de uma polaridade, a depender de quem se refere.

## 5.4 Análise de sentimento

Com a validação do modelo, foi iniciado a análise de sentimento das demais notícias dos dois indivíduos escolhidos para a pesquisa. Os resultados de cada análise de sentimento da notícia foi armazenado no banco para geração dos gráficos e visualização dos dados.

O autor do trabalho observou também uma grande quantidade de notícias do segundo indivíduo (Renan Calheiros) durante o mês de Dezembro de 2016. Este pico de notícias têm-se conta devido à quantidade de escândalos políticos durante esse período, como o afastamento e em seguida o retorno dele à presidência do Senado. O mesmo acontece para o primeiro indivíduo (Dilma Rousseff), que possuiu um pico de notícias após o início do ano de 2010, onde iniciou o primeiro mandato e outro na metade do ano de 2016, durante o período de *impeachment*. A [Tabela 4](#), mostra a quantidade de notícias divulgadas pelos canais de notícias mensalmente, a partir do mês de Outubro de 2016. Os dados divulgados são das notícias que conseguiram ser extraídas, excluindo notícias de blogs, vídeos e opiniões.

	10/2016	11/2016	12/2016	01/2017
Dilma Rousseff	69	43	60	58
Renan Calheiros	32	31	133	24

Tabela 4 – Quantidade de notícias em quatro meses

Como houveram problemas para capturar as informações dos outros canais de notícias, com exceção da Carta Capital, foi criado um gráfico mostrando a polaridade e a quantidade de notícias a partir de Janeiro de 2016 até Janeiro de 2017.

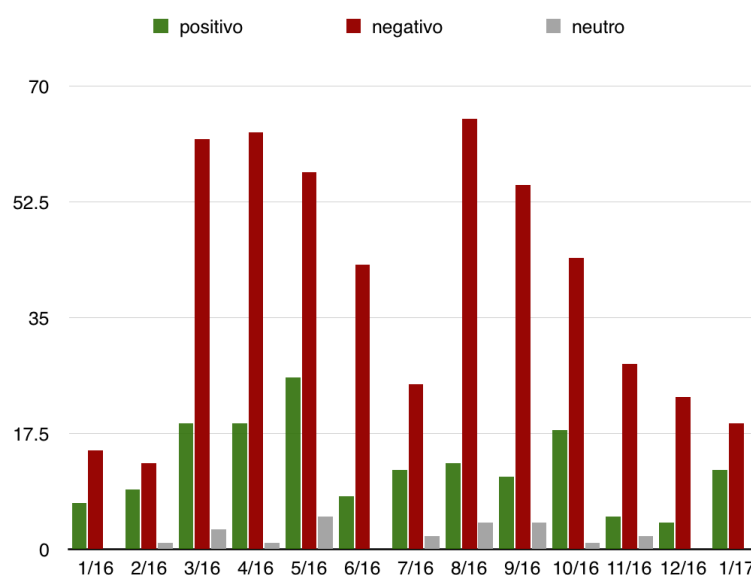


Figura 10 – Análise de sentimento da Carta Capital sobre Dilma Rousseff

Na Figura 10, é possível observar o período de *impeachment* da ex-presidente, durante o mês de Março de 2016 e Agosto de 2016. Nesse período houveram muitas notícias com sentimento negativo, como retrata o seguinte trecho da mídia: "Quase metade dos entrevistados que declararam voto em Dilma avaliam gestão como ruim ou péssima (...)"Carta Capital, 30/03/2016 e "Protestos na avenida Paulista, em São Paulo: pressão sobre Dilma aumenta (...)"Carta Capital, 14/03.2016. Entretanto, outro ponto chama atenção é a aparência de notícias positivas durante o período de *impeachment* da ex-presidente. As notícias negativas estavam relacionadas diretamente com a divulgação dos protestos e votação para *impeachment* de Dilma Rousseff.

Apesar de mostrar uma grande quantidade de notícias negativas no gráfico, é preciso reconhecer uma margem de erro para o algoritmo pois notícias que possuem como trecho: "Manter aquecido o mercado interno, estimular investimentos privados em infraestrutura e consolidar grandes grupos econômicos nacionais, sem reduzir gastos sociais e sem abrir mão da iniciativa pública em setores estratégicos, era uma combinação complicada de objetivos. Mas não inviável. (...)"Carta Capital, 06/01/2016. Esta notícia foi classificada como negativa, porém, deveria ser classificada como positiva. Este erro se justifica por existir uma negação positiva, em que é afirmado uma qualidade negando uma afirmação. Infelizmente o algoritmo não é capaz de tratar palavras ambíguas ou tais afirmações.

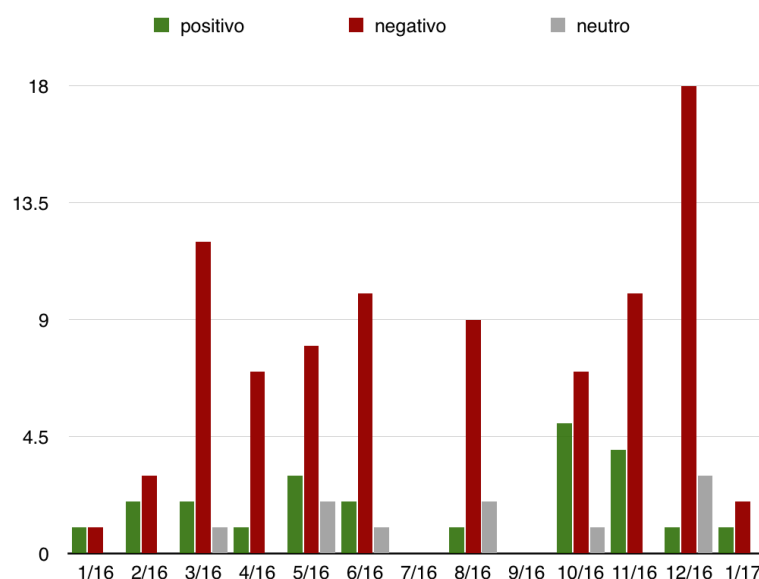


Figura 11 – Análise de sentimento da Carta Capital sobre Renan Calheiros

Na [Figura 11](#), trata-se da análise sentimental das notícias relacionadas à Renan Calheiros pela Carta Capital no mesmo período da [Figura 10](#). Durante o mesmo período do processo de *impeachment* da ex-presidente, começaram a surgir notícias negativas, como por exemplo: "O presidente do Senado, Renan Calheiros (PMDB-AL), anunciou na tarde desta segunda-feira (9) que está mantida para a próxima quarta-feira (11) a votação no plenário do Senado sobre o pedido de abertura do processo de impeachment contra Dilma Rousseff. (...)", UOL 09/05/2016. Outro período que pode-se chamar atenção é o de Dezembro de 2016, durante os escândalos de corrupção no Senado. Durante esse período é possível observar um grande volume de notícias negativas, como por exemplo: "O procurador-geral da República, Rodrigo Janot, apresentou nesta segunda-feira 12 ao Supremo Tribunal Federal (STF) uma denúncia de envolvimento do presidente do Senado, Renan Calheiros (PMDB-AL), em crimes investigados pela Operação Lava Jato. (...)"Carta Capital, 12/12/2016.

Portanto, observando as notícias da Carta Capital durante o período de um ano, é possível visualizar a mudança de comportamento e a quantidade de notícias positivas ou negativas quando existe algum fato importante. Entretanto, a determinação de saber se uma mídia é imparcial ou não vai além de apenas uma análise sentimental das notícias, pois é preciso realizar uma análise mais profunda, como uma comparação entre notícias do mesmo tipo e separação de fatos e opiniões. Para determinar o tipo

de uma notícia, é preciso realizar uma outra classificação para determinar notícias políticas, de esporte, economia e além disso, pode-se utilizar clusterização de notícias e então realizar a análise sentimental. Utilizando dessa maneira é possível realizar uma melhor comparação entre notícias, pois elas possuirão tipos similares.

A fim de validar os testes realizados e o conjunto de treinamento. Foram classificados 118 notícias manualmente pelo autor, mas elas não foram utilizadas para treinamento ou testes iniciais. A Figura 12 mostra o percentual de acerto e de erro do algoritmo com as 118 notícias, em que houveram 51 acertos e 67 erros de classificação de notícias para as classes de positiva, neutra e negativa.

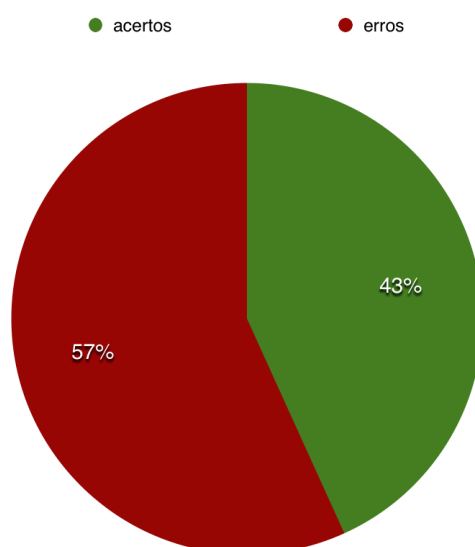


Figura 12 – Comparativo de acerto

Apesar da baixa taxa de acerto, o classificador conseguiu atingir um resultado razoável para classificação em três classes e ainda é possível melhorar esse resultado utilizando outras técnicas, aumentando a base de treinamento e aumentando a diversidade de notícias para treino e teste.

## 6 Conclusão

O trabalho apresentado mostrou como o processo de Mineração de Dados foi utilizado para coletar textos de notícias, processar e criar modelos de classificadores que realizam a análise sentimental, classificando os textos em classes pré-determinadas, que foram: positivas, negativas e neutras, e possibilitou a visualização do comportamento de um canal de notícia durante um período de tempo. Foi criada uma ferramenta escrita em Python para classificação automatizada de textos de notícias, onde os textos coletados estão disponíveis abertamente ao público.

Entretanto, o modelo de classificação gerado nesse trabalho não foi capaz de mostrar a imparcialidade dos principais canais de notícia brasileiros, devido à baixa taxa de acerto e ser um problema bem mais complexo, como explicado no Capítulo 5. Porém, a taxa de acerto foi suficiente para que pudesse refletir o sentimento geral das notícias durante certos períodos dos indivíduos de estudo e foi validado com os acontecimentos, como o impeachment de Dilma Rousseff e os escândalos no Senado por Renan Calheiros.

Os resultados obtidos trouxeram resultados que não eram esperados, como a baixa taxa de acerto, mostrando que a quantidade de notícias para treinamento não foi suficiente. Um outro fator mostrou-se com o estudo e coleta das notícias, como o foco da mídia altera em relação a um indivíduo, por diversos fatores, como escândalos ou por demanda da população. Para melhor observação e melhores conclusões sobre isso, é preciso realizar uma coleta de dados mais extensa sobre a mídia.

Apesar dos resultados negativos, foi mostrado nesse trabalho diferentes momentos do cenário político brasileiro, durante o período de Janeiro de 2016 e 2017, com relação aos dois indivíduos utilizados para estudo e a relação da Carta Capital com tais momentos. A ferramenta possibilitou a observação e retirada de pequenas conclusões de como um específico canal de notícia se comporta, apesar da baixa quantidade de dados extraídos.

Por fim, a ferramenta possui limitações de obtenção de texto dos artigos de notícia online, em que podem estar bloqueados para captura através do scrapy, como páginas que utilizam fortemente do *javascript* para mostrar o conteúdo ou realização de busca. Outra limitação encontra-se na própria página de busca do canal de notícia, em que pode haver problemas na paginação, onde de 30 páginas só é possível acessar até a décima, e outro exemplo é a limitação das notícias, em que são mostradas notícias apenas em um certo período. Existe também a limitação na classificação, em que a predição de sentimento não é garantida 100% e mesmo mostrando uma ótima taxa de acerto, ela está exposta a erros.

A fim de melhorar a comparação e o estudo, é importante também classificar as notícias com relação ao tipo conteúdo e dividi-las em categorias ou subcategorias para realizar a clusterização. Em seguida, serão obtidas notícias com categorias similares, possibilitando uma melhor comparação de sentimento, pois serão comparadas as notícias de economia com economia, o que seria diferente de comparar notícias de economia com pesquisa de voto.

Como conclusão, a ferramenta mostrou que é possível realizar a análise de sentimento de textos de notícias, porém precisa de alguns ajustes que são descritos na próxima seção. Em relação à pergunta feita no título, ela ainda encontra-se sem resposta, pois é preciso mais estudos e melhores comparações entre os textos de notícias.

## 6.1 Trabalhos Futuros

Em ordem de buscar melhores resultados, o modelo de classificação pode ser melhorado. Utilizando um realimentação do conjunto de treinamento com novas notícias e exemplos mais diversificados para que seja abrangido novos conjuntos de palavras. O uso de novas classes também pode ajudar no melhoramento do modelo.

Além disso, existe uma etapa para o pré-processamento onde as palavras de características são obtidas através do valor semântico da palavra na frase. Ao contrário de como a ferramenta atual trabalha, a escolha das palavras seriam filtradas apenas os adjetivos ou advérbios. Isso é possível com uma etapa de aprendizado semântico



no modelo através do aprendizado por padrões.

Existem outras técnicas de PLN, nesse trabalho foi utilizado apenas uma delas. Existem outras técnicas que podem ser empregadas, como o uso de matriz TF-IDF (*term frequency–inverse document frequency*), onde é criada uma matriz com os valores para indicar a importância de uma palavra no texto. A frequência do termo é dado pelo número de vezes que ele ocorre no texto, enquanto que o inverso da frequência no documento é dado para obter um peso maior nas palavras que aparecem menos no texto, e um peso menor para as palavras mais frequentes. A matriz TF-IDF é extremamente útil para ajudar a encontrar palavras de características nos textos positivos ou negativos, podendo ser utilizado para a descoberta de sentimento textual.

Por fim, os experimentos realizados e os textos obtidos foram restritos a apenas alguns canais de notícia, não havendo uma grande cobertura. Houve a limitação da busca, entretanto só foi realizada a procura por apenas uma fonte. Outras fontes de busca podem ser utilizadas para aumentar a quantidade de notícias de determinada mídia. Além disso, outras formas de captura de notícia diferente do scrapy podem ser utilizadas para suprir as limitações do *framework*.

# Anexos

# ANEXO A – Descrição de Requisitos

## A.1 Requisitos Funcionais

### A.1.1 Captura

#### [RF01] - Busca por indivíduo:

**Descrição:** Permite que a *spider* realize a busca do indivíduo na página do portal de notícia designado.

**Prioridade:** Alta.

**Saída:** Lista de notícias relacionadas ao indivíduo.

#### [RF02] - Obter texto da notícia:

**Descrição:** Permite que a *spider* acesse o artigo de notícia e obtenha o conteúdo do texto.

**Prioridade:** Alta.

**Saída:** O conteúdo textual da notícia é armazenado em memória.

#### [RF03] - Tratamento de texto e padronização:

**Descrição:** Permite que a *spider* remova elementos textuais que não fazem parte da notícia, como tags. Além disso, é realizado a padronização de itens, como a data, para permanecer em um formato unico.

**Prioridade:** Média.

**Saída:** O texto da notícia é filtrado para seguir um padrão.

#### [RF04] - Persistência de notícia:

**Descrição:** Permite que a *spider* persista uma notícia no banco de dados.

**Prioridade:** Alta.

**Saída:** Uma notícia inserida no banco de dados.

**[RF05] - Armazenar dados da captura:**

**Descrição:** Permite que a *spider* insira dados da captura de notícia durante a captura. Os dados são como: data de captura e termo buscado.

**Prioridade:** Média.

**Saída:** Os dados da captura da notícia são inseridos no banco de dados.

### A.1.2 Mineração de Texto

**[RF01] - Realizar pré-processamento de texto:**

**Descrição:** Permite realizar métodos de PLN para padronização de texto de notícia recebido.

**Prioridade:** Alta.

**Saída:** Uma lista de palavras em um formato padrão utilizado pela ferramenta.

**[RF02] - Treinamento do classificador:**

**Descrição:** Realização de treinamento do classificador utilizando uma base de dados já classificada.

**Prioridade:** Alta.

**Saída:** Modelo de classificador já treinado pela base.

**[RF03] - Predição da classificação do texto:**

**Descrição:** Permite a classificação de um texto através do algoritmo.

**Prioridade:** Alta.

**Saída:** O texto irá possuir uma das classes utilizadas pela ferramenta.

### A.1.3 API

#### [RF01] - Método GET para obter informações sobre um indivíduo:

**Descrição:** Permite que seja enviado uma mensagem à ferramenta utilizando o método GET para que seja retornado informações sobre o indivíduo informado na mensagem.

**Prioridade:** Média.

**Saída:** Uma mensagem em formato JSON com os dados das notícias relacionadas ao indivíduo.

#### [RF01] - Método POST para realizar a procura de indivíduo:

**Descrição:** Permite que seja enviado uma mensagem à ferramenta utilizando o método POST para que sejam agendadas execuções de busca de notícias do indivíduo informado na mensagem.

**Prioridade:** Baixa.

## A.2 Requisitos Não Funcionais

### A.2.1 [RNF01] Segurança dos dados

**Descrição:** É importante que exista uma preocupação e medidas para proteção dos dados obtidos para a possibilidade de perda ou alteração por terceiros.

**Prioridade:** Alta.

### A.2.2 [RNF02] Desempenho

**Descrição:** Embora não seja um requisito essencial ao sistema, medir o desempenho deve ser considerado por corresponder a um fator de qualidade de software.

**Prioridade:** Média.

### A.2.3 [RNF03] Disponibilidade do serviço

**Descrição:** Por ser um serviço que irá ser utilizado pela internet, é preciso que tenha uma taxa de disponibilidade superior a 90% do tempo, por exemplo.

**Prioridade:** Alta.

## ANEXO B – Notícias de Teste

Tabela 5 – Tabela de notícias negativas

Título	URL
Conceito de intervenção internacional de Dilma é criticado por especialistas	<a href="http://www.cartacapital.com.br/internacional/conceito-de-intervencao-internacional-de-dilma-e-criticado-por-especialistas-4409.html">http://www.cartacapital.com.br/internacional/conceito-de-intervencao-internacional-de-dilma-e-criticado-por-especialistas-4409.html</a>
Investigação da Lava Jato tem pontos fora da curva, diz Dilma Rousseff	<a href="http://www1.folha.uol.com.br/poder/2016/01/1732752-investigacao-da-lava-jato-tem-pontos-fora-da-curva-diz-dilma-rousseff.shtml">http://www1.folha.uol.com.br/poder/2016/01/1732752-investigacao-da-lava-jato-tem-pontos-fora-da-curva-diz-dilma-rousseff.shtml</a>
Em retaliação ao PT, Cunha acolhe pedido de impeachment contra Dilma	<a href="http://www.cartacapital.com.br/politica/em-retaliacao-ao-pt-cunha-acolhe-pedido-de-impeachment-contradilma-5647.html">http://www.cartacapital.com.br/politica/em-retaliacao-ao-pt-cunha-acolhe-pedido-de-impeachment-contradilma-5647.html</a>
CNT / MDA: avaliação positiva do governo é de 7,7%	<a href="http://www.cartacapital.com.br/blogs/parlatorio/cnt-mda-avaliacao-positiva-do-governo-e-de-7-7-6507.html">http://www.cartacapital.com.br/blogs/parlatorio/cnt-mda-avaliacao-positiva-do-governo-e-de-7-7-6507.html</a>
Rejeição de Serra sobe e de Dilma cai, diz Sensus	<a href="http://www.cartacapital.com.br/politica/rejeicao-de-serra-sobe-e-de-dilma-cai-diz-sensus">http://www.cartacapital.com.br/politica/rejeicao-de-serra-sobe-e-de-dilma-cai-diz-sensus</a>
Governo Dilma é ruim ou péssimo para 69% dos brasileiros, diz Ibope	<a href="http://www.cartacapital.com.br/blogs/parlatorio/69-consideram-governo-dilma-ruim-ou-pessimo-diz-ibope">http://www.cartacapital.com.br/blogs/parlatorio/69-consideram-governo-dilma-ruim-ou-pessimo-diz-ibope</a>
Paulo Henrique Amorim obtém vitória no TSE	<a href="http://www.cartacapital.com.br/politica/paulo-henrique-amorim-obtem-vitoria-no-tse">http://www.cartacapital.com.br/politica/paulo-henrique-amorim-obtem-vitoria-no-tse</a>
Não existe onda verde, diz Guedes, do Sensus	<a href="http://www.cartacapital.com.br/politica/nao-existe-onda-verde-diz-guedes-do-sensus">http://www.cartacapital.com.br/politica/nao-existe-onda-verde-diz-guedes-do-sensus</a>
Dilma oferece mais R\$ 701 milhões para financiar porto cubano	<a href="http://www1.folha.uol.com.br/mundo/2014/01/1403485-dilma-oferece-mais-r-701-milhoes-para-financiar-porto-cubano.shtml">http://www1.folha.uol.com.br/mundo/2014/01/1403485-dilma-oferece-mais-r-701-milhoes-para-financiar-porto-cubano.shtml</a>

**Tabela 5 - Tabela de notícias negativas (continuação da página anterior)**

Título	URL
Torcedores xingam Dilma Rousseff e gritam 'fora PT' antes de jogo do Brasil	<a href="http://www1.folha.uol.com.br/esporte/2016/03/1754231-torcedores-xingam-dilma-rousseff-e-gritam-fora-pt-antes-de-jogo-do-brasil.shtml">http://www1.folha.uol.com.br/esporte/2016/03/1754231-torcedores-xingam-dilma-rousseff-e-gritam-fora-pt-antes-de-jogo-do-brasil.shtml</a>
Dilma se trumbica - Tantas foram as mentiras de Dilma sobre a Petrobras que o maior rombo a cobrir agora é o da credibilidade	<a href="http://epoca.globo.com/colunas-e-blogs/ruth-de-aquino/noticia/2015/01/dilma-bse-trumbicab.html">http://epoca.globo.com/colunas-e-blogs/ruth-de-aquino/noticia/2015/01/dilma-bse-trumbicab.html</a>
Dilma acertou: a oposição precisa estudar - A sucção do Estado pelo PT é tão sofisticada que exige décadas de estudo para ser compreendida	<a href="http://epoca.globo.com/colunas-e-blogs/guilherme-fiuza/noticia/2013/11/bdilma-acertou-b-oposicao-precisa-estudar.html">http://epoca.globo.com/colunas-e-blogs/guilherme-fiuza/noticia/2013/11/bdilma-acertou-b-oposicao-precisa-estudar.html</a>
Dilma foi citada 11 vezes nos depoimentos de delatores da Lava Jato	<a href="http://www1.folha.uol.com.br/poder/2015/03/1601387-dilma-foi-citada-11-vezes-nos-depoimentos-de-delatores-da-lava-jato.shtml">http://www1.folha.uol.com.br/poder/2015/03/1601387-dilma-foi-citada-11-vezes-nos-depoimentos-de-delatores-da-lava-jato.shtml</a>
Dilma Rousseff: da ascensão à queda - Ostentando os maiores índices de popularidade de um presidente eleito desde Fernando Collor, Dilma viu o jogo virar após as manifestações de 2013 e acabou cometendo erros que levaram a seu impeachment	<a href="http://epoca.globo.com/tempo/especial-impeachment/noticia/2016/08/dilma-rousseff-da-ascensao-quedaa.html">http://epoca.globo.com/tempo/especial-impeachment/noticia/2016/08/dilma-rousseff-da-ascensao-quedaa.html</a>
Dilma diz que Uber tira emprego de muitas pessoas - Mesmo assim, presidente adotou uma postura ponderada sobre a regulamentação. "Não é a União que decide", disse	<a href="http://epoca.globo.com/vida/experiencias-digitais/noticia/2015/09/dilma-diz-que-uber-tira-emprego-de-muitas-pessoas.html">http://epoca.globo.com/vida/experiencias-digitais/noticia/2015/09/dilma-diz-que-uber-tira-emprego-de-muitas-pessoas.html</a>



**Tabela 5 - Tabela de notícias negativas (continuação da página anterior)**

Título	URL
“Dilma não precisava atravessar o continente para falar que vai cumprir a lei” - ONGs criticam a declaração da presidente sobre acabar com desmatamento ilegal na Amazônia, o que já está na lei	<a href="http://epoca.globo.com/colunas-e-blogs/blog-do-planeta/amazonia/noticia/2015/06/dilma-nao-precisava-atravesar-o-continente-para-falar-que-vai-cumprir-lei.html">http://epoca.globo.com/colunas-e-blogs/blog-do-planeta/amazonia/noticia/2015/06/dilma-nao-precisava-atravesar-o-continente-para-falar-que-vai-cumprir-lei.html</a>
Países bolivarianos reagem a cassação de Dilma Rousseff e condenam 'golpe'	<a href="http://www1.folha.uol.com.br/mundo/2016/08/1808755-evo-diz-que-ira-chamar-embaixador-da-bolivia-se-dilma-sofrer-impeachment.shtml">http://www1.folha.uol.com.br/mundo/2016/08/1808755-evo-diz-que-ira-chamar-embaixador-da-bolivia-se-dilma-sofrer-impeachment.shtml</a>
Veja argumentos usados por quem quer o impeachment	<a href="http://noticias.uol.com.br/politica/ultimas-noticias/2015/03/12/veja-argumentos-usados-por-quem-quer-impeachment.htm">http://noticias.uol.com.br/politica/ultimas-noticias/2015/03/12/veja-argumentos-usados-por-quem-quer-impeachment.htm</a>
PMDB oficializa sua saída do governo Dilma em reunião de poucos minutos	<a href="http://noticias.uol.com.br/politica/ultimas-noticias/2016/03/29/pmdb-oficializa-sua-saida-do-governo-dilma-em-reuniao-de-poucos-minutos.htm">http://noticias.uol.com.br/politica/ultimas-noticias/2016/03/29/pmdb-oficializa-sua-saida-do-governo-dilma-em-reuniao-de-poucos-minutos.htm</a>
Diante da crise, uma reforma política radical	<a href="http://www.cartacapital.com.br/blogs/outras-palavras/diante-da-crise-uma-reforma-politica-radical">http://www.cartacapital.com.br/blogs/outras-palavras/diante-da-crise-uma-reforma-politica-radical</a>
Aécio some em debate polarizado por Dilma e Marina	<a href="http://www.cartacapital.com.br/politica/dilma-e-marina-silva-polarizam-debate-do-sbt-2831.html">http://www.cartacapital.com.br/politica/dilma-e-marina-silva-polarizam-debate-do-sbt-2831.html</a>
Jucá articulou desembarque do PMDB do governo Dilma Rousseff	<a href="http://www1.folha.uol.com.br/poder/2016/05/1770848-juca-articulou-desembarque-do-pmdb-do-governo-dilma-rousseff.shtml">http://www1.folha.uol.com.br/poder/2016/05/1770848-juca-articulou-desembarque-do-pmdb-do-governo-dilma-rousseff.shtml</a>
Edinho Silva vai prestar depoimento no processo contra Dilma e Temer no TSE	<a href="http://oglobo.globo.com/brasil/edinho-silva-vai-prestar-depoimento-no-processo-contradilma-temer-no-tse-20377908">http://oglobo.globo.com/brasil/edinho-silva-vai-prestar-depoimento-no-processo-contradilma-temer-no-tse-20377908</a>

**Tabela 5 - Tabela de notícias negativas (continuação da página anterior)**

<b>Título</b>	<b>URL</b>
Teori devolve ação da Lava Jato contra Lula ao STF	< <a href="http://www.cartacapital.com.br/blogs/parlatorio/acao-da-lava-jato-contralula-volta-ao-stf-entenda">http://www.cartacapital.com.br/blogs/parlatorio/acao-da-lava-jato-contralula-volta-ao-stf-entenda</a> >
Grampo que fez Gilmar anular nomeação de Lula é ilegal, diz Teori	< <a href="http://www.cartacapital.com.br/blogs/parlatorio/grampo-que-fez-gilmar-anular-nomeacao-de-lula-e-ilegal-diz-teori">http://www.cartacapital.com.br/blogs/parlatorio/grampo-que-fez-gilmar-anular-nomeacao-de-lula-e-ilegal-diz-teori</a> >
Dilma diz que não autorizou caixa 2 em campanha	< <a href="http://www1.folha.uol.com.br/poder/2016/07/1794274-dilma-diz-que-nao-autorizou-caixa-2-em-campanha-delacoes-preocupam-pt.shtml">http://www1.folha.uol.com.br/poder/2016/07/1794274-dilma-diz-que-nao-autorizou-caixa-2-em-campanha-delacoes-preocupam-pt.shtml</a> >
Delcídio admite propina de Pasadena e levanta suspeita sobre Dilma	< <a href="http://www1.folha.uol.com.br/poder/2016/03/1750123-delcidio-admite-propina-de-pasadena-e-levanta-suspeita-sobre-dilma.shtml">http://www1.folha.uol.com.br/poder/2016/03/1750123-delcidio-admite-propina-de-pasadena-e-levanta-suspeita-sobre-dilma.shtml</a> >
Conheça os inquilinos dos imóveis de Dilma Rousseff	< <a href="http://www1.folha.uol.com.br/poder/2014/08/1495045-conheca-os-inquilinos-dos-imoveis-de-dilma-rousseff.shtml">http://www1.folha.uol.com.br/poder/2014/08/1495045-conheca-os-inquilinos-dos-imoveis-de-dilma-rousseff.shtml</a> >
Calçada de apresentador é pichada: “Morra Jô Soares”	< <a href="http://www.cartacapital.com.br/blogs/parlatorio/calçada-de-apresentador-e-pichada-201cmorra-jo-soares201d-8801.html">http://www.cartacapital.com.br/blogs/parlatorio/calçada-de-apresentador-e-pichada-201cmorra-jo-soares201d-8801.html</a> >
Brasil vai apoiar todas as medidas adotadas pela ONU em relação à Líbia, diz porta-voz	< <a href="http://www.cartacapital.com.br/internacional/brasil-vai-apoiar-todas-as-medidas-adotadas-pela-onu-em-relacao-a-libia-diz-porta-voz">http://www.cartacapital.com.br/internacional/brasil-vai-apoiar-todas-as-medidas-adotadas-pela-onu-em-relacao-a-libia-diz-porta-voz</a> >
“Eu nunca disse que (o dinheiro) era para o PT”, afirma consultor à PF	< <a href="http://www.cartacapital.com.br/politica/-e2-80-9ceu-nunca-disse-que-o-dinheiro-era-para-o-pt-e2-80-9d-afirma-consultor-a-pf">http://www.cartacapital.com.br/politica/-e2-80-9ceu-nunca-disse-que-o-dinheiro-era-para-o-pt-e2-80-9d-afirma-consultor-a-pf</a> >
“Lula é o farol do movimento sindical e da política mundial”	< <a href="http://www.cartacapital.com.br/internacional/lula-e-o-farol-do-movimento-sindical-e-da-politica-mundial">http://www.cartacapital.com.br/internacional/lula-e-o-farol-do-movimento-sindical-e-da-politica-mundial</a> >

**Tabela 5 - Tabela de notícias negativas (continuação da página anterior)**

<b>Título</b>	<b>URL</b>
Crise de Dilma é mais grave que a de Collor no pré-impeachment, diz Ronaldo Caiado	< <a href="http://noticias.uol.com.br/politica/ultimas-noticias/2014/12/08/crise-de-dilma-e-mais-grave-que-a-de-collor-no-pre-impeachment-diz-ronaldo-caiado.htm">http://noticias.uol.com.br/politica/ultimas-noticias/2014/12/08/crise-de-dilma-e-mais-grave-que-a-de-collor-no-pre-impeachment-diz-ronaldo-caiado.htm</a> >
Em nota, CNA anuncia apoio a impeachment de Dilma Rousseff	< <a href="http://www1.folha.uol.com.br/poder/2016/04/1758022-em-nota-cna-anuncia-apoio-a-impeachment-de-dilma-rousseff.shtml">http://www1.folha.uol.com.br/poder/2016/04/1758022-em-nota-cna-anuncia-apoio-a-impeachment-de-dilma-rousseff.shtml</a> >
Imprensa europeia vê "insurreição de hipócritas" no impeachment	< <a href="http://www.cartacapital.com.br/internacional/imprensa-europeia-ve-insurreicao-de-hipocritas-no-impeachment">http://www.cartacapital.com.br/internacional/imprensa-europeia-ve-insurreicao-de-hipocritas-no-impeachment</a> >
Aécio Neves: impeachment não está na agenda do PSDB	< <a href="http://www.cartacapital.com.br/blogs/parlatorio/impeachment-nao-esta-na-agenda-do-psdb-diz-aecio-neves-7730.html">http://www.cartacapital.com.br/blogs/parlatorio/impeachment-nao-esta-na-agenda-do-psdb-diz-aecio-neves-7730.html</a> >
C, de cangaço	< <a href="http://www.cartacapital.com.br/revista/841/c-de-cangaco-817.html">http://www.cartacapital.com.br/revista/841/c-de-cangaco-817.html</a> >
Artigo da presidente Dilma Rousseff é epitáfio de seu governo, diz leitor	< <a href="http://www1.folha.uol.com.br/paineldoleitor/2016/04/1761734-artigo-da-presidente-dilma-rousseff-e-epitafio-de-seu-governo-diz-leitor.shtml">http://www1.folha.uol.com.br/paineldoleitor/2016/04/1761734-artigo-da-presidente-dilma-rousseff-e-epitafio-de-seu-governo-diz-leitor.shtml</a> >
Como o mundo está vendo o impeachment de Dilma Rousseff	< <a href="http://economia.uol.com.br/noticias/infomoney/2016/08/31/como-o-mundo-esta-vendo-o-impeachment-de-dilma-rousseff.htm">http://economia.uol.com.br/noticias/infomoney/2016/08/31/como-o-mundo-esta-vendo-o-impeachment-de-dilma-rousseff.htm</a> >
De como o PT derrubou a si mesmo - A retórica de vitimização de Dilma e Lula vai começar a carecer de sentido. O fôlego dessa propaganda é curto	< <a href="http://epoca.globo.com/colunas-e-blogs/eugenio-bucci/noticia/2016/05/de-como-o-pt-derrubou-si-mesmo.html">http://epoca.globo.com/colunas-e-blogs/eugenio-bucci/noticia/2016/05/de-como-o-pt-derrubou-si-mesmo.html</a> >
Leitores comentam o impeachment da presidente Dilma Rousseff	< <a href="http://www1.folha.uol.com.br/paineldoleitor/2016/09/1809063-leitores-comentam-o-impeachment-da-presidente-dilma-rousseff.shtml">http://www1.folha.uol.com.br/paineldoleitor/2016/09/1809063-leitores-comentam-o-impeachment-da-presidente-dilma-rousseff.shtml</a> >

**Tabela 5 - Tabela de notícias negativas (continuação da página anterior)**

Título	URL
A língua das imagens	< <a href="http://www.cartacapital.com.br/politica/a-lingua-das-imagens">http://www.cartacapital.com.br/politica/a-lingua-das-imagens</a> >
Teori nega pedido de Dilma para suspender impeachment	< <a href="http://oglobo.globo.com/brasil/teori-nega-pedido-de-dilma-para-suspender-impeachment-20325564">http://oglobo.globo.com/brasil/teori-nega-pedido-de-dilma-para-suspender-impeachment-20325564</a> >
Temer vai substituir programa de fronteira criado pelo governo Dilma	< <a href="http://oglobo.globo.com/brasil/temer-vai-substituir-programa-de-fronteira-criado-pelo-governo-dilma-20472164">http://oglobo.globo.com/brasil/temer-vai-substituir-programa-de-fronteira-criado-pelo-governo-dilma-20472164</a> >
Aécio leva denúncia da Veja ao seu último horário eleitoral	< <a href="http://www.cartacapital.com.br/blogs/cartas-eleicoes/aecio-leva-denuncia-da-veja-em-seu-ultimo-horario-eleitoral-7045.html">http://www.cartacapital.com.br/blogs/cartas-eleicoes/aecio-leva-denuncia-da-veja-em-seu-ultimo-horario-eleitoral-7045.html</a> >
Ao anunciar novas vagas, Dilma se enrola para explicar meta do Pronatec	< <a href="http://www1.folha.uol.com.br/educacao/2015/07/1661345-ao-anunciar-novas-vagas-dilma-se-enrola-para-explicar-meta-do-pronatec.shtml">http://www1.folha.uol.com.br/educacao/2015/07/1661345-ao-anunciar-novas-vagas-dilma-se-enrola-para-explicar-meta-do-pronatec.shtml</a> >
Ex-presidente da Andrade Gutierrez muda depoimento e nega propina para a chapa Dilma-Temer	< <a href="http://oglobo.globo.com/brasil/ex-presidente-da-andrade-gutierrez-muda-depoimento-nega-propina-para-chapa-dilma-temer-20484863">http://oglobo.globo.com/brasil/ex-presidente-da-andrade-gutierrez-muda-depoimento-nega-propina-para-chapa-dilma-temer-20484863</a> >
ONU Mulheres condena violência sexista praticada contra Dilma	< <a href="http://www.cartacapital.com.br/politica/onu-mulheres-condena-violencia-sexista-praticada-contradilma">http://www.cartacapital.com.br/politica/onu-mulheres-condena-violencia-sexista-praticada-contradilma</a> >
Em discurso no Senado, Dilma diz temer a morte da democracia	< <a href="http://www1.folha.uol.com.br/poder/2016/08/1808274-em-discurso-no-senado-dilma-diz-temer-a-morte-da-democracia.shtml">http://www1.folha.uol.com.br/poder/2016/08/1808274-em-discurso-no-senado-dilma-diz-temer-a-morte-da-democracia.shtml</a> >
Onde está Wally, quer dizer, Dilma Rousseff? - Dilma pode ser uma “carta fora do baralho”. Mas o jogo ainda tem lances emocionantes pela frente	< <a href="http://epoca.globo.com/colunas-e-blogs/eugenio-bucci/noticia/2016/06/onde-esta-wally-quer-dizer-dilma-rousseff.html">http://epoca.globo.com/colunas-e-blogs/eugenio-bucci/noticia/2016/06/onde-esta-wally-quer-dizer-dilma-rousseff.html</a> >

**Tabela 5 - Tabela de notícias negativas (continuação da página anterior)**

Título	URL
Governo Dilma é desaprovado por 69% e aprovado por 10%, diz Ibope	<a href="http://noticias.uol.com.br/politica/ultimas-noticias/2016/03/30/69-consideram-gestao-de-dilma-ruim-ou-pessima-aponta-pesquisa-ibope.htm">http://noticias.uol.com.br/politica/ultimas-noticias/2016/03/30/69-consideram-gestao-de-dilma-ruim-ou-pessima-aponta-pesquisa-ibope.htm</a>
Caiu o único ministro de Dilma - O programa do PT foi saudado com um pannelaço. Faltou João Santana com sua interpretação dos fatos	<a href="http://epoca.globo.com/colunas-e-blogs/guilherme-fiuza/noticia/2016/03/caiu-o-unico-ministro-de-dilma.html">http://epoca.globo.com/colunas-e-blogs/guilherme-fiuza/noticia/2016/03/caiu-o-unico-ministro-de-dilma.html</a>
Lava Jato coloca campanhas do PT no alvo	<a href="http://www.cartacapital.com.br/blogs/parlatorio/lava-jato-coloca-campanhas-do-pt-no-alvo">http://www.cartacapital.com.br/blogs/parlatorio/lava-jato-coloca-campanhas-do-pt-no-alvo</a>
STF nega liminares para anular fatiamento do julgamento de Dilma	<a href="http://www1.folha.uol.com.br/poder/2016/09/1812069-stf-nega-liminares-para-anular-fatiamento-do-julgamento-de-dilma.shtml">http://www1.folha.uol.com.br/poder/2016/09/1812069-stf-nega-liminares-para-anular-fatiamento-do-julgamento-de-dilma.shtml</a>
Mudança de Dilma a Porto Alegre custou R\$ 60 mil para governo federal	<a href="http://www1.folha.uol.com.br/poder/2016/09/1810681-mudanca-de-dilma-a-porto-alegre-custou-r-60-mil-para-governo-temer.shtml">http://www1.folha.uol.com.br/poder/2016/09/1810681-mudanca-de-dilma-a-porto-alegre-custou-r-60-mil-para-governo-temer.shtml</a>
Dilma Rousseff foi espionada pelos EUA	<a href="http://www.cartacapital.com.br/internacional/dilma-rousseff-foi-espionada-pelos-eua-6006.html">http://www.cartacapital.com.br/internacional/dilma-rousseff-foi-espionada-pelos-eua-6006.html</a>
TCU deve seguir relator e rejeitar contas de Dilma Rousseff de 2015	<a href="http://oglobo.globo.com/brasil/tcu-deve-seguir-relator-rejeitar-contas-de-dilma-rousseff-de-2015-20237107">http://oglobo.globo.com/brasil/tcu-deve-seguir-relator-rejeitar-contas-de-dilma-rousseff-de-2015-20237107</a>
Manifestações de domingo podem aumentar pressão pelo impeachment de Dilma	<a href="http://noticias.uol.com.br/politica/ultimas-noticias/2016/03/12/manifestacoes-de-domingo-podem-aumentar-pressao-pelo-impeachment-de-dilma.htm">http://noticias.uol.com.br/politica/ultimas-noticias/2016/03/12/manifestacoes-de-domingo-podem-aumentar-pressao-pelo-impeachment-de-dilma.htm</a>
Dilma está pendurada no Supremo - Não é fácil desmontar um presépio de pobres coitados que têm um cheque especial do tamanho da Petrobras	<a href="http://epoca.globo.com/colunas-e-blogs/guilherme-fiuza/noticia/2016/03/dilma-esta-pendurada-no-supremo.html">http://epoca.globo.com/colunas-e-blogs/guilherme-fiuza/noticia/2016/03/dilma-esta-pendurada-no-supremo.html</a>

**Tabela 5 - Tabela de notícias negativas (continuação da página anterior)**

Título	URL
Cerveró, use a máscara de Dilma - É errado dizer que o PT está pagando seus pecados. Quem está pagando os pecados do governo petista são os contribuintes	<a href="http://epoca.globo.com/colunas-e-blogs/guilherme-fiuza/noticia/2015/07/cervero-use-mascara-de-dilma.html">http://epoca.globo.com/colunas-e-blogs/guilherme-fiuza/noticia/2015/07/cervero-use-mascara-de-dilma.html</a>
Gays, negros e eleitores de Dilma - Como é se sentir parte de uma minoria segregada em sua própria cidade	<a href="http://epoca.globo.com/colunas-e-blogs/ivan-martins/noticia/2014/10/gays-negros-e-eleitores-de-dilma.html">http://epoca.globo.com/colunas-e-blogs/ivan-martins/noticia/2014/10/gays-negros-e-eleitores-de-dilma.html</a>
Reprovação do governo chega a 64%, diz CNI/Ibope	<a href="http://www.cartacapital.com.br/blogs/parlatorio/reprovacao-do-governo-chega-a-64-diz-cni-ibope-9385.html">http://www.cartacapital.com.br/blogs/parlatorio/reprovacao-do-governo-chega-a-64-diz-cni-ibope-9385.html</a>
Balanço da Petrobras complica Dilma - Dilma é instrumento de um projeto imoral de poder. O país não tem o direito de linchá-la, mas tem o dever de investigá-la	<a href="http://epoca.globo.com/colunas-e-blogs/guilherme-fiuza/noticia/2015/06/balanco-da-petrobras-complica-dilma.html">http://epoca.globo.com/colunas-e-blogs/guilherme-fiuza/noticia/2015/06/balanco-da-petrobras-complica-dilma.html</a>
Depois de relação próxima, Dilma diz que Cabral 'jamais' foi aliado	<a href="http://oglobo.globo.com/brasil/depois-de-relacao-proxima-dilma-diz-que-cabral-jamais-foi-aliado-20482595">http://oglobo.globo.com/brasil/depois-de-relacao-proxima-dilma-diz-que-cabral-jamais-foi-aliado-20482595</a>
Dilma Rousseff sanciona MP do Futebol	<a href="http://www1.folha.uol.com.br/esporte/2015/08/1664851-dilma-rousseff-sanciona-mp-do-futebol.shtml">http://www1.folha.uol.com.br/esporte/2015/08/1664851-dilma-rousseff-sanciona-mp-do-futebol.shtml</a>
Dilma Rousseff: "As instituições no Brasil são mais sólidas do que se imagina- Na primeira entrevista após ser afastada da presidência, Dilma diz ao jornalista Glenn Greenwald que o Brasil não vive o "fim da democracia"	<a href="http://epoca.globo.com/colunas-e-blogs/ruth-de-aquino/noticia/2016/05/dilma-rousseff-instituicoes-no-brasil-sao-mais-solidas-do-que-se-imagina.html">http://epoca.globo.com/colunas-e-blogs/ruth-de-aquino/noticia/2016/05/dilma-rousseff-instituicoes-no-brasil-sao-mais-solidas-do-que-se-imagina.html</a>

**Tabela 5 - Tabela de notícias negativas (continuação da página anterior)**

<b>Título</b>	<b>URL</b>
PSB e Rede se posicionam a favor do impeachment de Dilma Rousseff	< <a href="http://www1.folha.uol.com.br/poder/2016/04/1759915-psb-e-rede-se-posicionam-a-favor-do-impeachment-de-dilma-rousseff.shtml">http://www1.folha.uol.com.br/poder/2016/04/1759915-psb-e-rede-se-posicionam-a-favor-do-impeachment-de-dilma-rousseff.shtml</a> >
Dilma chama Temer e Cunha de 'chefes conspiradores' em discurso no Planalto	< <a href="http://noticias.uol.com.br/politica/ultimas-noticias/2016/04/12/dilma-encontro-da-educacao.htm">http://noticias.uol.com.br/politica/ultimas-noticias/2016/04/12/dilma-encontro-da-educacao.htm</a> >
Após ter mandato cassado, Dilma Rousseff mantém os direitos políticos	< <a href="http://noticias.uol.com.br/politica/ultimas-noticias/2016/08/31/apos-ter-mandato-cassado-dilma-rousseff-mantem-os-direitos-politicos.htm">http://noticias.uol.com.br/politica/ultimas-noticias/2016/08/31/apos-ter-mandato-cassado-dilma-rousseff-mantem-os-direitos-politicos.htm</a> >
"Não queria destruir a presidente Dilma- Após a prisão, João Santana admite caixa dois de campanha. "Só não sabia que era dinheiro sujo"	< <a href="http://epoca.globo.com/colunas-e-blogs/ruth-de-aquino/noticia/2016/07/nao-queria-destruir-presidente-dilma.html">http://epoca.globo.com/colunas-e-blogs/ruth-de-aquino/noticia/2016/07/nao-queria-destruir-presidente-dilma.html</a> >
Ocaso do governo Dilma instaura clima de fim de expediente em Brasília	< <a href="http://www1.folha.uol.com.br/poder/2016/05/1766551-ocaso-do-governo-dilma-instaura-clima-de-fim-de-expediente-em-brasilia.shtml">http://www1.folha.uol.com.br/poder/2016/05/1766551-ocaso-do-governo-dilma-instaura-clima-de-fim-de-expediente-em-brasilia.shtml</a> >
Dilma diz que não sente 'ódio' de Temer, mas o chama de 'traidor'	< <a href="http://www1.folha.uol.com.br/poder/2016/10/1822475-dilma-diz-que-nao-sente-odio-de-temer-mas-o-chama-de-traidor.shtml">http://www1.folha.uol.com.br/poder/2016/10/1822475-dilma-diz-que-nao-sente-odio-de-temer-mas-o-chama-de-traidor.shtml</a> >
Conversa entre Dilma e Lula é grampeada pela PF	< <a href="http://www.cartacapital.com.br/blogs/parlatorio/conversa-entre-dilma-e-lula-e-grampeada-pela-pf">http://www.cartacapital.com.br/blogs/parlatorio/conversa-entre-dilma-e-lula-e-grampeada-pela-pf</a> >
Romero Jucá nega saída da liderança do governo	< <a href="http://www.cartacapital.com.br/politica/romero-juca-nega-saida-da-lideranca-do-governo">http://www.cartacapital.com.br/politica/romero-juca-nega-saida-da-lideranca-do-governo</a> >
STF quebra sigilo do DEM e de Agripino Maia	< <a href="http://www.cartacapital.com.br/politica/supremo-quebra-sigilo-do-dem-e-de-agripino-maia">http://www.cartacapital.com.br/politica/supremo-quebra-sigilo-do-dem-e-de-agripino-maia</a> >
A popularidade de... FHC	< <a href="http://www.cartacapital.com.br/revista/870/a-popularidade-de-fhc-4315.html">http://www.cartacapital.com.br/revista/870/a-popularidade-de-fhc-4315.html</a> >

**Tabela 5 - Tabela de notícias negativas (continuação da página anterior)**

Título	URL
Grampos, Lula, Dilma e Moro: entenda o novo caos político	<a href="http://www.cartacapital.com.br/politica/grampos-lula-dilma-e-moro-entenda-o-novo-caos-politico">http://www.cartacapital.com.br/politica/grampos-lula-dilma-e-moro-entenda-o-novo-caos-politico</a>
Lula: "FHC deveria ficar quieto"	<a href="http://www.cartacapital.com.br/politica/lula-fhc-deveria-ficar-quieto">http://www.cartacapital.com.br/politica/lula-fhc-deveria-ficar-quieto</a>
STF não deve avaliar hoje pedido de abertura do processo de Dilma durante a ditadura	<a href="http://www.cartacapital.com.br/politica/stf-nao-deve-avaliar-hoje-pedido-de-abertura-do-processo-de-dilma-durante-a-ditadura">http://www.cartacapital.com.br/politica/stf-nao-deve-avaliar-hoje-pedido-de-abertura-do-processo-de-dilma-durante-a-ditadura</a>
Ex-marido de Dilma fala sobre atuação na guerrilha e roubo do cofre do Adhemar	<a href="http://noticias.uol.com.br/politica/ultimas-noticias/2011/06/29/ex-marido-de-dilma-fala-sobre-atuacao-na-guerrilha-e-roubo-do-cofre-do-adhemar.htm">http://noticias.uol.com.br/politica/ultimas-noticias/2011/06/29/ex-marido-de-dilma-fala-sobre-atuacao-na-guerrilha-e-roubo-do-cofre-do-adhemar.htm</a>
A armadilha da popularidade do governo Dilma	<a href="http://www.cartacapital.com.br/economia/a-armadilha-da-popularidade-do-governo-dilma">http://www.cartacapital.com.br/economia/a-armadilha-da-popularidade-do-governo-dilma</a>
Programa de Serra usa "caso Erenice" e Dirceu para atacar Dilma	<a href="http://www.cartacapital.com.br/politica/programa-de-serra-usa-caso-erenice-e-dirceu-para-atacar-dilma">http://www.cartacapital.com.br/politica/programa-de-serra-usa-caso-erenice-e-dirceu-para-atacar-dilma</a>
Avaliação de Dilma cai 20 pontos desde junho	<a href="http://www.cartacapital.com.br/politica/avaliacao-de-dilma-cai-20-pontos-desde-junho-7874.html">http://www.cartacapital.com.br/politica/avaliacao-de-dilma-cai-20-pontos-desde-junho-7874.html</a>
Após mais de 20 horas, Senado aprova processo de impeachment e afasta Dilma	<a href="http://noticias.uol.com.br/politica/ultimas-noticias/2016/05/12/senado-aprova-processo-de-impeachment-e-afasta-dilma-por-ate-180-dias.htm">http://noticias.uol.com.br/politica/ultimas-noticias/2016/05/12/senado-aprova-processo-de-impeachment-e-afasta-dilma-por-ate-180-dias.htm</a>
PT busca Lula, e Dilma deve viver exílio interno	<a href="http://www1.folha.uol.com.br/poder/2016/05/1770881-pt-busca-lula-e-dilma-deve-viver-exilio-interno.shtml">http://www1.folha.uol.com.br/poder/2016/05/1770881-pt-busca-lula-e-dilma-deve-viver-exilio-interno.shtml</a>



**Tabela 5 - Tabela de notícias negativas (continuação da página anterior)**

<b>Título</b>	<b>URL</b>
Dilma não cumpriu 43% das promessas de 2010	< <a href="http://www1.folha.uol.com.br/poder/2014/09/1524018-dilma-nao-cumpriu-43-das-promessas-de-2010-confira-a-lista-das-69-feitas.shtml">http://www1.folha.uol.com.br/poder/2014/09/1524018-dilma-nao-cumpriu-43-das-promessas-de-2010-confira-a-lista-das-69-feitas.shtml</a> >
Dilma, a solitária - Qual naufraga perdida numa ilha, ela lança garrafas sobre as ondas e não acha destinatário	< <a href="http://epoca.globo.com/colunas-e-blogs/eugenio-bucci/noticia/2013/07/dilma-b-solitariab.html">http://epoca.globo.com/colunas-e-blogs/eugenio-bucci/noticia/2013/07/dilma-b-solitariab.html</a> >
Gleisi diz que Senado não tem "moral" para julgar Dilma e gera bate-boca	< <a href="http://noticias.uol.com.br/politica/ultimas-noticias/2016/08/25/gleisi-diz-que-senado-nao-tem-moral-para-julgar-dilma-e-gera-bate-boca.htm">http://noticias.uol.com.br/politica/ultimas-noticias/2016/08/25/gleisi-diz-que-senado-nao-tem-moral-para-julgar-dilma-e-gera-bate-boca.htm</a> >
Senado vota e passa à fase final do processo de impeachment de Dilma	< <a href="http://www1.folha.uol.com.br/poder/2016/08/1801051-senado-torna-dilma-re-e-votacao-indica-aprovacao-do-impeachment.shtml">http://www1.folha.uol.com.br/poder/2016/08/1801051-senado-torna-dilma-re-e-votacao-indica-aprovacao-do-impeachment.shtml</a> >
O primeiro golpe democrático da história - Ninguém jamais resistiu com tanta fibra ao ridículo. Dilma conseguiu tornar-se golpista de si mesma	< <a href="http://epoca.globo.com/colunas-e-blogs/guilherme-fiuza/noticia/2016/08/o-primeiro-golpe-democratico-da-historia.html">http://epoca.globo.com/colunas-e-blogs/guilherme-fiuza/noticia/2016/08/o-primeiro-golpe-democratico-da-historia.html</a> >

**Tabela 6 – Tabela de notícias positivas**

<b>Título</b>	<b>URL</b>
Dilma fará pronunciamento sobre aprovação do impeachment na Câmara	< <a href="http://www1.folha.uol.com.br/poder/2016/04/1762332-dilma-fara-pronunciamento-sobre-aprovacao-do-impeachment-na-camara.shtml">http://www1.folha.uol.com.br/poder/2016/04/1762332-dilma-fara-pronunciamento-sobre-aprovacao-do-impeachment-na-camara.shtml</a> >
No Parlamento alemão, impeachment de Dilma é chamado de golpe e conspiração	< <a href="http://www.cartacapital.com.br/politica/no-parlamento-alemao-impeachment-de-dilma-e-chamado-de-golpe-e-conspiracao">http://www.cartacapital.com.br/politica/no-parlamento-alemao-impeachment-de-dilma-e-chamado-de-golpe-e-conspiracao</a> >

**Tabela 6 - Tabela de notícias positivas (continuação da página anterior)**

<b>Título</b>	<b>URL</b>
A militante Dilma e os arquivos	< <a href="http://www.cartacapital.com.br/politica/a-militante-dilma-e-os-arquivos">http://www.cartacapital.com.br/politica/a-militante-dilma-e-os-arquivos</a> >
Dilma considera resultado positivo	< <a href="http://www.cartacapital.com.br/sustentabilidade/dilma-considera-resultado-positivo">http://www.cartacapital.com.br/sustentabilidade/dilma-considera-resultado-positivo</a> >
Interrompido por bate-boca, Cardozo volta a defender Dilma na Câmara	< <a href="http://noticias.uol.com.br/politica/ultimas-noticias/2016/04/11/interrompido-por-bate-boca-cardozo-volta-a-defender-dilma-na-camara.htm">http://noticias.uol.com.br/politica/ultimas-noticias/2016/04/11/interrompido-por-bate-boca-cardozo-volta-a-defender-dilma-na-camara.htm</a> >
Juristas vão a Brasília dar apoio a Dilma Rousseff	< <a href="http://www1.folha.uol.com.br/colunas/monicabergamo/2016/03/1752365-juristas-vaao-brasilia-dar-apoio-a-dilma-rousseff.shtml">http://www1.folha.uol.com.br/colunas/monicabergamo/2016/03/1752365-juristas-vaao-brasilia-dar-apoio-a-dilma-rousseff.shtml</a> >
Dilma Rousseff é a segunda mulher mais poderosa do mundo na lista da Forbes	< <a href="http://noticias.uol.com.br/ultimas-noticias/afp/2013/05/22/dilma-rousseff-e-a-segunda-mulher-mais-poderosa-do-mundo-na-lista-da-forbes.htm">http://noticias.uol.com.br/ultimas-noticias/afp/2013/05/22/dilma-rousseff-e-a-segunda-mulher-mais-poderosa-do-mundo-na-lista-da-forbes.htm</a> >
Dilma diz que apoiará plebiscito e pede para que Senado a inocente	< <a href="http://noticias.uol.com.br/politica/ultimas-noticias/2016/08/16/dilma-divulga-mensagem-ao-senado-e-ao-povo-brasileiro.htm">http://noticias.uol.com.br/politica/ultimas-noticias/2016/08/16/dilma-divulga-mensagem-ao-senado-e-ao-povo-brasileiro.htm</a> >
FHC, o hipócrita monumental	< <a href="http://www.cartacapital.com.br/revista/905/fhc-o-hipocrita-monumental">http://www.cartacapital.com.br/revista/905/fhc-o-hipocrita-monumental</a> >
Dilma contesta 'conclusões genéricas' do MPE sobre sua chapa com Temer em 2014	< <a href="http://oglobo.globo.com/brasil/dilma-contesta-conclusoes-genericas-do-mpe-sobre-sua-chapa-com-temer-em-2014-20659286">oglobo.globo.com/brasil/dilma-contesta-conclusoes-genericas-do-mpe-sobre-sua-chapa-com-temer-em-2014-20659286</a> >
Dilma: "A gritaria dos golpistas não vai me tirar do rumo"	< <a href="http://www.cartacapital.com.br/blogs/parlatorio/dilma-gritaria-dos-golpistas-nao-vai-me-tirar-do-rumo">http://www.cartacapital.com.br/blogs/parlatorio/dilma-gritaria-dos-golpistas-nao-vai-me-tirar-do-rumo</a> >
O que explica a futura vitória de Dilma Rousseff?	< <a href="http://www.cartacapital.com.br/economia/o-que-explica-a-futura-vitoria-de-dilma-rousseff">http://www.cartacapital.com.br/economia/o-que-explica-a-futura-vitoria-de-dilma-rousseff</a> >

**Tabela 6 - Tabela de notícias positivas (continuação da página anterior)**

<b>Título</b>	<b>URL</b>
Até o papa apoiou Serra	< <a href="http://www.cartacapital.com.br/politica/ate-o-papa-apoia-serra">http://www.cartacapital.com.br/politica/ate-o-papa-apoia-serra</a> >
Lula e Dilma discutiram a Lava Jato em Brasília	< <a href="http://www.cartacapital.com.br/blogs/cartas-da-esplanada/lula-e-dilma-discutiram-a-lava-jato-em-brasilia-933.html">http://www.cartacapital.com.br/blogs/cartas-da-esplanada/lula-e-dilma-discutiram-a-lava-jato-em-brasilia-933.html</a> >
Crise foi transformada em oportunidade no País	< <a href="http://www.cartacapital.com.br/politica/brasil-transforma-momento-de-criese-em-oportunidade-avalia-dilma">http://www.cartacapital.com.br/politica/brasil-transforma-momento-de-criese-em-oportunidade-avalia-dilma</a> >
"Democracia é acatar o voto da maioria", diz Lula na Paulista	< <a href="http://www.cartacapital.com.br/politica/democracia-e-acatar-o-voto-da-maioria-diz-lula-na-paulista">http://www.cartacapital.com.br/politica/democracia-e-acatar-o-voto-da-maioria-diz-lula-na-paulista</a> >
Datafolha mostra pequena recuperação da presidente	< <a href="http://www1.folha.uol.com.br/poder/2015/12/1721228-datafolha-mostra-pequena-recuperacao-da-presidente.shtml">http://www1.folha.uol.com.br/poder/2015/12/1721228-datafolha-mostra-pequena-recuperacao-da-presidente.shtml</a> >
Dilma apoia "plebiscito" por Constituinte e reforma política	< <a href="http://www.cartacapital.com.br/blogs/cartas-eleicoes/dilma-anuncia-apoio-a-plebiscito-popular-por-constituente-exclusiva-e-reforma-politica-6961.html">http://www.cartacapital.com.br/blogs/cartas-eleicoes/dilma-anuncia-apoio-a-plebiscito-popular-por-constituente-exclusiva-e-reforma-politica-6961.html</a> >
Inflação não subirá no Brasil, afirma Mantega	< <a href="http://www.cartacapital.com.br/politica/inflacao-nao-subira-no-brasil-afirma-mantega">http://www.cartacapital.com.br/politica/inflacao-nao-subira-no-brasil-afirma-mantega</a> >
Dilma afirma que "o Brasil inteiro está #FechadoComOTinga"	< <a href="http://www.cartacapital.com.br/sociedade/dilma-afirma-que-o-brasil-inteiro-esta-fechadocomotinga-1517.html">http://www.cartacapital.com.br/sociedade/dilma-afirma-que-o-brasil-inteiro-esta-fechadocomotinga-1517.html</a> >
Tracking Vox Populi aponta vitória de Dilma Rousseff no primeiro turno	< <a href="http://www.cartacapital.com.br/politica/tracking-vox-populi-aponta-vitoria-de-dilma-rousseff-no-primeiro-turno">http://www.cartacapital.com.br/politica/tracking-vox-populi-aponta-vitoria-de-dilma-rousseff-no-primeiro-turno</a> >
Dilma Rousseff presidenta do Brasil	< <a href="http://www.cartacapital.com.br/politica/dilma-rousseff-presidenta-do-brasil">http://www.cartacapital.com.br/politica/dilma-rousseff-presidenta-do-brasil</a> >

**Tabela 6 - Tabela de notícias positivas (continuação da página anterior)**

<b>Título</b>	<b>URL</b>
Chico Buarque assina manifesto em apoio a Dilma Rousseff	< <a href="http://www.cartacapital.com.br/politica/chico-buarque-assina-manifesto-em-apoio-a-dilma-rousseff">http://www.cartacapital.com.br/politica/chico-buarque-assina-manifesto-em-apoio-a-dilma-rousseff</a> >
Obama, Putin, Merkel e Mujica parabenizam Dilma	< <a href="http://www.cartacapital.com.br/politica/obama-putin-e-outros-lideres-mundiais-parabenizam-dilma-433.html">http://www.cartacapital.com.br/politica/obama-putin-e-outros-lideres-mundiais-parabenizam-dilma-433.html</a> >
Lula termina governo com aprovação de 83,4%, diz CNT/Sensus	< <a href="http://www.cartacapital.com.br/politica/lula-termina-governo-com-aprovacao-de-834-diz-cntsensus">http://www.cartacapital.com.br/politica/lula-termina-governo-com-aprovacao-de-834-diz-cntsensus</a> >
Governo federal pretende conceder 75 mil bolsas de estudo no exterior até 2014	< <a href="http://www.cartacapital.com.br/politica/governo-vai-conceder-75-mil-bolsas-de-estudo-no-externo-ate-2014">http://www.cartacapital.com.br/politica/governo-vai-conceder-75-mil-bolsas-de-estudo-no-externo-ate-2014</a> >
Senado aprova Luiz Fachin, indicado por Dilma, para o STF	< <a href="http://www.cartacapital.com.br/blogs/parlatorio/senado-aprova-luiz-fachin-indicado-por-dilma-para-o-stf-5284.html">http://www.cartacapital.com.br/blogs/parlatorio/senado-aprova-luiz-fachin-indicado-por-dilma-para-o-stf-5284.html</a> >
Dilma se exalta com pergunta sobre "incompetência" feita por jornalista da Al Jazeera	< <a href="http://noticias.uol.com.br/politica/ultimas-noticias/2016/12/16/dilma-se-exalta-com-pergunta-sobre-incompetencia-feita-por-jornalista-da-al-jazeera.htm">http://noticias.uol.com.br/politica/ultimas-noticias/2016/12/16/dilma-se-exalta-com-pergunta-sobre-incompetencia-feita-por-jornalista-da-al-jazeera.htm</a> >
Dilma e sua faxina	< <a href="http://www.cartacapital.com.br/politica/dilma-e-sua-faxina">http://www.cartacapital.com.br/politica/dilma-e-sua-faxina</a> >
Dilma é Dilma	< <a href="http://www.cartacapital.com.br/politica/dilma-e-dilma">http://www.cartacapital.com.br/politica/dilma-e-dilma</a> >
Dilma: "Não renuncio em hipótese alguma"	< <a href="http://www.cartacapital.com.br/blogs/parlatorio/dilma-diz-que-nao-renunciara-e-que-impeachment-e-tentativa-de-golpe">http://www.cartacapital.com.br/blogs/parlatorio/dilma-diz-que-nao-renunciara-e-que-impeachment-e-tentativa-de-golpe</a> >
Iriny Lopes deixa Secretaria de Políticas para as Mulheres	< <a href="http://www.cartacapital.com.br/politica/iriny-lopes-deixa-secretaria-de-politicas-para-as-mulheres">http://www.cartacapital.com.br/politica/iriny-lopes-deixa-secretaria-de-politicas-para-as-mulheres</a> >

**Tabela 6 - Tabela de notícias positivas (continuação da página anterior)**

<b>Título</b>	<b>URL</b>
De Portugal, Dilma e Lula falam sobre José Alencar	< <a href="http://www.cartacapital.com.br/politica/planalto-divulga-nota-de-pesar-pela-morte-de-jose-alencar">http://www.cartacapital.com.br/politica/planalto-divulga-nota-de-pesar-pela-morte-de-jose-alencar</a> >
Dilma confirma ausência na festa de aniversário do PT	< <a href="http://www1.folha.uol.com.br/poder/2016/02/1744167-dilma-confirma-ausencia-na-festa-de-aniversario-do-pt.shtml">http://www1.folha.uol.com.br/poder/2016/02/1744167-dilma-confirma-ausencia-na-festa-de-aniversario-do-pt.shtml</a> >
Dilma diz que comentário de Gilmar Mendes foi inadequado	< <a href="http://oglobo.globo.com/brasil/dilma-diz-que-comentario-de-gilmar-mendes-foi-inadequado-20434859">http://oglobo.globo.com/brasil/dilma-diz-que-comentario-de-gilmar-mendes-foi-inadequado-20434859</a> >
Admiradora de Dilma, jovem candidata a vereadora no RS adota sobrenome Rousseff	< <a href="http://noticias.uol.com.br/ultimas-noticias/agencia-estado/2016/08/23/admiradora-de-dilma-jovem-candidata-a-vereadora-no-rs-adota-sobrenome-rousseff.htm">http://noticias.uol.com.br/ultimas-noticias/agencia-estado/2016/08/23/admiradora-de-dilma-jovem-candidata-a-vereadora-no-rs-adota-sobrenome-rousseff.htm</a> >
Banco do Brasil diz ao Senado que pedaladas do Plano Safra não têm ato de Dilma	< <a href="http://noticias.uol.com.br/politica/ultimas-noticias/2016/06/17/banco-do-brasil-diz-ao-senado-que-pedaladas-do-plano-safra-nao-tem-ato-de-dilma.htm">http://noticias.uol.com.br/politica/ultimas-noticias/2016/06/17/banco-do-brasil-diz-ao-senado-que-pedaladas-do-plano-safra-nao-tem-ato-de-dilma.htm</a> >
Dilma venceria todos os concorrentes no 2º turno	< <a href="http://www.cartacapital.com.br/politica/dilma-venceria-todos-os-concorrentes-no-2o-turno-7543.html">http://www.cartacapital.com.br/politica/dilma-venceria-todos-os-concorrentes-no-2o-turno-7543.html</a> >
Dilma tem oito pontos de vantagem no primeiro turno e empata com Marina no segundo	< <a href="http://www.cartacapital.com.br/politica/dilma-tem-oito-pontos-de-vantagem-no-primeiro-turno-e-empata-com-marina-no-segundo-7754.html">http://www.cartacapital.com.br/politica/dilma-tem-oito-pontos-de-vantagem-no-primeiro-turno-e-empata-com-marina-no-segundo-7754.html</a> >
Após receber intimação, Dilma diz que cometeu 'erros', mas não 'crimes'	< <a href="http://www1.folha.uol.com.br/poder/2016/05/1770526-apos-receber-intimacao-dilma-diz-que-cometeu-erros-mas-nao-crimes.shtml">http://www1.folha.uol.com.br/poder/2016/05/1770526-apos-receber-intimacao-dilma-diz-que-cometeu-erros-mas-nao-crimes.shtml</a> >
Argentina inaugura política externa de Dilma	< <a href="http://www.cartacapital.com.br/politica/argentina-inaugura-politica-externa-de-dilma">http://www.cartacapital.com.br/politica/argentina-inaugura-politica-externa-de-dilma</a> >

**Tabela 6 - Tabela de notícias positivas (continuação da página anterior)**

<b>Título</b>	<b>URL</b>
Obama elogia Dilma e diz que relação entre países pode aumentar	< <a href="http://www.cartacapital.com.br/internacional/obama-elogia-dilma-e-diz-que-relacao-entre-paises-pode-aumentar">http://www.cartacapital.com.br/internacional/obama-elogia-dilma-e-diz-que-relacao-entre-paises-pode-aumentar</a> >
Dez anos da Lei Maria da Penha: nenhum retrocesso é aceitável	< <a href="http://noticias.uol.com.br/opiniao/coluna/2016/08/07/dez-anos-da-lei-maria-da-penha-nenhum-retrocesso-e-aceitavel.htm">http://noticias.uol.com.br/opiniao/coluna/2016/08/07/dez-anos-da-lei-maria-da-penha-nenhum-retrocesso-e-aceitavel.htm</a> >
Dólar cai com sinais de saída do PMDB da base do governo Dilma Rousseff	< <a href="http://www1.folha.uol.com.br/mercado/2016/03/1755006-dolar-cai-com-sinais-de-saida-do-pmdb-da-base-do-governo-dilma-rousseff.shtml">http://www1.folha.uol.com.br/mercado/2016/03/1755006-dolar-cai-com-sinais-de-saida-do-pmdb-da-base-do-governo-dilma-rousseff.shtml</a> >
As certezas férreas de Dilma Rousseff	< <a href="http://www.cartacapital.com.br/politica/as-certezas-ferreas-de-dilma-rousseff-8021.html">http://www.cartacapital.com.br/politica/as-certezas-ferreas-de-dilma-rousseff-8021.html</a> >
Dilma Rousseff - Primeira mulher a ser eleita presidente no Brasil, Dilma sofreu impeachment no Congresso. Porém, conseguiu manter o direito de assumir cargo público	< <a href="http://epoca.globo.com/tudo-sobre/noticia/2016/06/dilma-rousseff.html">http://epoca.globo.com/tudo-sobre/noticia/2016/06/dilma-rousseff.html</a> >
Vencendo ou perdendo, PT seguirá influente	< <a href="http://www.cartacapital.com.br/politica/vencendo-ou-perdendo-pt-seguira-influente-182.html">http://www.cartacapital.com.br/politica/vencendo-ou-perdendo-pt-seguira-influente-182.html</a> >
Lula diz estar de volta à política, em congresso	< <a href="http://www.cartacapital.com.br/politica/lula-diz-estar-de-volta-a-politica-em-congresso-de-sindicalistas">http://www.cartacapital.com.br/politica/lula-diz-estar-de-volta-a-politica-em-congresso-de-sindicalistas</a> >
Dilma, mostre que é de briga	< <a href="http://www.cartacapital.com.br/politica/dilma-mostre-que-e-de-briga">http://www.cartacapital.com.br/politica/dilma-mostre-que-e-de-briga</a> >
O simbólico apoio do Brasil a Evo Morales	< <a href="http://www.cartacapital.com.br/internacional/o-simbolico-apoio-do-brasil-a-evo-morales">http://www.cartacapital.com.br/internacional/o-simbolico-apoio-do-brasil-a-evo-morales</a> >
Dilma retoma fôlego	< <a href="http://www.cartacapital.com.br/revista/764/dilma-retoma-folego-4622.html">http://www.cartacapital.com.br/revista/764/dilma-retoma-folego-4622.html</a> >

**Tabela 6 - Tabela de notícias positivas (continuação da página anterior)**

<b>Título</b>	<b>URL</b>
Dilma pede desculpas por andar de carro com o neto no colo	< <a href="http://www.cartacapital.com.br/politica/dilma-pede-desculpas-por-andar-de-carro-com-o-neto-no-colo-3865.html">http://www.cartacapital.com.br/politica/dilma-pede-desculpas-por-andar-de-carro-com-o-neto-no-colo-3865.html</a> >
Na pesquisa espontânea, Lula é o 2º nome mais citado	< <a href="http://www.cartacapital.com.br/politica/na-pesquisa-espontanea-lula-e-o-2o-nome-mais-citado-3850.html">http://www.cartacapital.com.br/politica/na-pesquisa-espontanea-lula-e-o-2o-nome-mais-citado-3850.html</a> >
Uma página virada no governo	< <a href="http://www.cartacapital.com.br/politica/uma-pagina-virada-no-governo">http://www.cartacapital.com.br/politica/uma-pagina-virada-no-governo</a> >
O padrão Olga Curado	< <a href="http://www.cartacapital.com.br/politica/o-padrao-olga-curado">http://www.cartacapital.com.br/politica/o-padrao-olga-curado</a> >
Dilma Rousseff condena redução da maioria penal	< <a href="http://www.cartacapital.com.br/blogs/parlatorio/dilma-rousseff-condena-reducao-da-maioridade-penal-7775.html">http://www.cartacapital.com.br/blogs/parlatorio/dilma-rousseff-condena-reducao-da-maioridade-penal-7775.html</a> >
Crise faz presidente Dilma Rousseff demonstrar irritação	< <a href="http://www1.folha.uol.com.br/poder/2015/07/1654560-cri-se-faz-presidente-dilma-rousseff-demonstrar-irritacao.shtml">http://www1.folha.uol.com.br/poder/2015/07/1654560-cri-se-faz-presidente-dilma-rousseff-demonstrar-irritacao.shtml</a> >
Aprovação do governo Dilma Rousseff sobe de 59% para 62%	< <a href="http://www.cartacapital.com.br/politica/aprovacao-do-governo-dilma-rousseff-sobe-de-59-para-62">http://www.cartacapital.com.br/politica/aprovacao-do-governo-dilma-rousseff-sobe-de-59-para-62</a> >
Governo Dilma é aprovado por 62% dos brasileiros	< <a href="http://www.cartacapital.com.br/politica/governo-dilma-e-aprovado-por-62-dos-brasileiros">http://www.cartacapital.com.br/politica/governo-dilma-e-aprovado-por-62-dos-brasileiros</a> >
É boato: Dilma não foi expulsa de pizzaria em São Paulo	< <a href="http://noticias.uol.com.br/politica/ultimas-noticias/2016/06/05/e-boato-dilma-nao-foi-expulsa-de-pizzaria-em-sao-paulo.htm">http://noticias.uol.com.br/politica/ultimas-noticias/2016/06/05/e-boato-dilma-nao-foi-expulsa-de-pizzaria-em-sao-paulo.htm</a> >
Os 4 institutos já divulgaram pesquisas para o 2º Turno	< <a href="http://www.cartacapital.com.br/politica/os-4-institutos-ja-divulgaram-pesquisas-para-o-2-c2-ba-turno">http://www.cartacapital.com.br/politica/os-4-institutos-ja-divulgaram-pesquisas-para-o-2-c2-ba-turno</a> >

Tabela 7 – Tabela de notícias neutras

Título	URL
Ação do PSDB no TSE: Cunha pode virar presidente do Brasil?	< <a href="http://www.cartacapital.com.br/politica/tse-reabre-acao-que-pede-a-cassacao-da-chapa-eleitoral-de-dilma-390.html">http://www.cartacapital.com.br/politica/tse-reabre-acao-que-pede-a-cassacao-da-chapa-eleitoral-de-dilma-390.html</a> >
Cid Gomes deixa o Ministério da Educação	< <a href="http://www.cartacapital.com.br/politica/cid-gomes-deixa-o-ministerio-da-educacao-7760.html">http://www.cartacapital.com.br/politica/cid-gomes-deixa-o-ministerio-da-educacao-7760.html</a> >
Dilma: "tenho muito respeito pelo ET de Varginha"	< <a href="http://www.cartacapital.com.br/sociedade/dilma-tenho-muito-respeito-pelo-et-de-varginha-8055.html">http://www.cartacapital.com.br/sociedade/dilma-tenho-muito-respeito-pelo-et-de-varginha-8055.html</a> >
Os principais pontos da política econômica de Temer e Meirelles	< <a href="http://www.cartacapital.com.br/economia/Os-principais-pontos-da-politica-economica-de-temer-e-meirelles">http://www.cartacapital.com.br/economia/Os-principais-pontos-da-politica-economica-de-temer-e-meirelles</a> >
Gilmar Mendes é acusado de crime de responsabilidade	< <a href="http://www.cartacapital.com.br/politica/gilmar-mendes-e-acusado-de-crime-de-responsabilidade">http://www.cartacapital.com.br/politica/gilmar-mendes-e-acusado-de-crime-de-responsabilidade</a> >
PT aprova indicação de Dilma para conselho de fundação	< <a href="http://oglobo.globo.com/brasil/pt-aprova-indicacao-de-dilma-para-conselho-de-fundacao-20448010">http://oglobo.globo.com/brasil/pt-aprova-indicacao-de-dilma-para-conselho-de-fundacao-20448010</a> >
Alexandre Garcia comenta o pronunciamento de Dilma Rousseff	< <a href="http://g1.globo.com/bom-dia-brasil/noticia/2016/08/alexandre-garcia-comenta-o-pronunciamento-de-dilma-rousseff.html">http://g1.globo.com/bom-dia-brasil/noticia/2016/08/alexandre-garcia-comenta-o-pronunciamento-de-dilma-rousseff.html</a> >
Brasília e a fome do poder	< <a href="http://www.cartacapital.com.br/revista/918/brasilia-e-a-fome-do-poder">http://www.cartacapital.com.br/revista/918/brasilia-e-a-fome-do-poder</a> >
Julgamento de Dilma pode demorar até sete dias	< <a href="http://www.cartacapital.com.br/politica/julgamento-de-dilma-deve-demorar-ate-sete-dias">http://www.cartacapital.com.br/politica/julgamento-de-dilma-deve-demorar-ate-sete-dias</a> >
Governo e oposição fecham acordo de procedimento para votação do Mínimo	< <a href="http://www.cartacapital.com.br/economia/governo-e-oposicao-fecham-acordo-de-procedimento-para-votacao-do-minimo">http://www.cartacapital.com.br/economia/governo-e-oposicao-fecham-acordo-de-procedimento-para-votacao-do-minimo</a> >



**Tabela 7 - Tabela de notícias neutras (continuação da página anterior)**

<b>Título</b>	<b>URL</b>
Dilma lamenta morte de ex-ministro	< <a href="http://www.cartacapital.com.br/politica/dilma-lamenta-morte-de-ex-ministro">http://www.cartacapital.com.br/politica/dilma-lamenta-morte-de-ex-ministro</a> >
Dilma e os Direitos Humanos	< <a href="http://www.cartacapital.com.br/politica/dilma-e-os-direitos-humanos">http://www.cartacapital.com.br/politica/dilma-e-os-direitos-humanos</a> >
Maradona manda mensagem de apoio a Dilma no Facebook. Presidente agradece - Mensagem foi enviada ontem (13/12), em mais um dia de manifestações de apoiadores do impeachment da presidente	< <a href="http://epoca.globo.com/vida/experiencias-digitais/noticia/2015/12/maradona-manda-mensagem-de-apoio-dilma-no-facebook-presidente-agradece.html">http://epoca.globo.com/vida/experiencias-digitais/noticia/2015/12/maradona-manda-mensagem-de-apoio-dilma-no-facebook-presidente-agradece.html</a> >
Defesa de Dilma Rousseff pede nova perícia de contas ao TSE	< <a href="http://oglobo.globo.com/brasil/defesa-de-dilma-rousseff-pede-nova-pericia-de-contas-ao-tse-20078378">http://oglobo.globo.com/brasil/defesa-de-dilma-rousseff-pede-nova-pericia-de-contas-ao-tse-20078378</a> >
TSE ameaça Temer no pós-impeachment	< <a href="http://www.cartacapital.com.br/politica/tse-ameaca-temer-no-pos-impeachment">http://www.cartacapital.com.br/politica/tse-ameaca-temer-no-pos-impeachment</a> >
Qual é o rito do impeachment?	< <a href="http://www.cartacapital.com.br/politica/qual-e-o-rito-do-impeachment-4435.html">http://www.cartacapital.com.br/politica/qual-e-o-rito-do-impeachment-4435.html</a> >
Destino, o fator privado	< <a href="http://www.cartacapital.com.br/destaques_carta_capital/destino-o-fator-privado">http://www.cartacapital.com.br/destaques_carta_capital/destino-o-fator-privado</a> >
Começou a defesa da Petrobras	< <a href="http://www.cartacapital.com.br/politica/comecou-a-defesa-da-petrobras-8687.html">http://www.cartacapital.com.br/politica/comecou-a-defesa-da-petrobras-8687.html</a> >
Marta oficializa sua saída da disputa por prefeitura de SP	< <a href="http://www.cartacapital.com.br/politica/marta-oficializa-sua-saida-da-disputa-por-prefeitura-de-sp">http://www.cartacapital.com.br/politica/marta-oficializa-sua-saida-da-disputa-por-prefeitura-de-sp</a> >
Entenda os próximos passos do impeachment contra Dilma Rousseff	< <a href="http://www.cartacapital.com.br/politica/entenda-os-proximos-passos-para-a-abertura-do-processo-de-impeachment">http://www.cartacapital.com.br/politica/entenda-os-proximos-passos-para-a-abertura-do-processo-de-impeachment</a> >

**Tabela 7 - Tabela de notícias neutras (continuação da página anterior)**

<b>Título</b>	<b>URL</b>
Aécio lidera entre eleitores mais escolarizados	<a href="http://www.cartacapital.com.br/politica/aecio-lidera-entre-eleitores-mais-escolarizados-8288.html">http://www.cartacapital.com.br/politica/aecio-lidera-entre-eleitores-mais-escolarizados-8288.html</a>

## Referências

- ANGELOV, P. P.; ZHOU, X. Evolving fuzzy-rule-based classifiers from data streams. *IEEE Transactions on Fuzzy Systems*, IEEE, p. 1462–1475, 2008. Citado na página 12.
- ARANHA, C. N. Uma abordagem de pré- processamento automático para mineração de textos em português: Sob o enfoque da inteligência computacional. *Tese de Doutorado - PUC-RJ*, 2007. Citado 5 vezes nas páginas 6, 18, 19, 26 e 27.
- BARBOSA, F. *Jornais apostam em sofisticação no impresso e expansão do digital*. 2015. Citado na página 11.
- CAMBRIA, E.; WHITE, B. Jumping NLP Curves: A Review of Natural Language Processing Research [Review Article]. *IEEE Computational Intelligence Magazine*, v. 9, n. 2, p. 48–57, maio 2014. ISSN 1556-603X. Disponível em: <<http://dx.doi.org/10.1109/mci.2014.2307227>>. Citado na página 19.
- CHICKERUR, S.; GOUDAR, A.; KINNERKAR, A. Comparison of relational database with document-oriented database (mongodb) for big data applications. *8th International Conference on Advanced Software Engineering*, IEEE, 2015. Citado 2 vezes nas páginas 24 e 25.
- CNNIC. *Statistical report on internet development in china*. 2013. Citado na página 11.
- CNNIC. *Internet users by country*. 2016. Disponível em: <<http://www.internetlivestats.com/internet-users-by-country/>>. Citado na página 11.
- DARSHAKAR, A. Crawler intelligence with machine learning and data mining integration. *ICPC*, IEEE, 2015. Citado na página 36.
- ESULI; SEBASTIANI. Sentwordnet: A publicly available lexical resource for opinion mining. *Conference on Language Resources and Evaluation (LREC)*, p. 417– 422, 2006. Citado na página 16.
- FARHOODI, M.; YARI, A.; SAYAH, A. N-gram based text classification for persian newspaper corpus. In: *The 7th International Conference on Digital Content, Multimedia Technology and its Applications*. [S.l.: s.n.], 2011. p. 55–59. Citado na página 16.
- FELDMAN, R.; DAGAN, I. Kdt - knowledge discovery from web news. *International Conference on Knowledge Discovery*, p. 112–117, 1995. Citado na página 12.
- FIELDING, R. T. *Architectural Styles and the Design of Network-based Software Architectures*. Tese (Doutorado), 2000. AAI9980887. Citado na página 16.
- FOUNDATION, P. S. *The Python Standard Library*. 2017. Disponível em: <<https://docs.python.org/3/library/asyncio.html>>. Citado na página 37.

- FOUNDATION, P. S. *The Python Tutorial*. 2017. Disponível em: <<https://docs.python.org/2/tutorial/datastructures.html#tuples-and-sequences>>. Citado na página 29.
- GOMES, H. J. C. Text mining: análise de sentimentos na classificação de notícias. *Information Systems and Technologies (CISTI), 2013 8th Iberian Conference*, 2013. Citado 2 vezes nas páginas 18 e 21.
- HSU, L. Mining on term extraction from web news. Springer, Heidelberg, p. 188–194, 1995. Citado na página 12.
- IGLESIAS, J. A. et al. Web news mining in an evolving framework. *IDEAL*, Springer, p. 327–335, 2014. Citado na página 12.
- INC., D. *What is Docker?* 2015. Disponível em: <<https://www.docker.com/what-docker>>. Citado na página 37.
- INDURKHIA, N.; DAMERAU, F. Handbook of natural language processing. *2ed Florida: CRC Press*, 2010. Citado na página 20.
- KANDT, R. K. Software requirements engineering: Practices and techniques. *Jet Propulsion Laboratory*, California Institute of Technology, 2003. Citado na página 33.
- KOTHAPALLI, M.; SHARIFAHMADIAN, E.; SHIH, L. Data mining of social media for analysis of product review. *International Journal of Computer Applications*, Foundation of Computer Science (FCS), 2016. Citado na página 25.
- KOULOUMPIS, E.; WILSON, T.; MOORE, J. Twitter Sentiment Analysis: The Good the Bad and the OMG! *Proceedings of the Fifth International AAI Conference on Weblogs and Social Media*, 2011. Disponível em: <<http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewFile/2857/3251>>. Citado na página 15.
- KOUZIS-LOUKAS, D. *Learning Scrapy*. [S.l.]: Packt Publishing, 2016. 648 p. Citado 2 vezes nas páginas 21 e 23.
- LEUNG, C. K.-S.; MACKINNON, R. K.; JIANG, F. Reducing the search space for big data mining for interesting patterns from uncertain data. *International Congress on Big Data*, IEEE, 2014. Citado na página 36.
- LIDDY, E. D. Natural language processing. In: *Encyclopedia of Library and Information Science*. 2nd Ed. NY: Marcel Decker, Inc., 2001. Citado 2 vezes nas páginas 19 e 20.
- LOPER, E. Nltk: Building a pedagogical toolkit in python. 2004. Citado na página 30.
- LOPER, E. Nltk: the natural language toolkit. *Second Annual PyCon Conference*, 2004. Citado na página 25.
- MAURO, A. D.; GRECO, M.; GRIMALDI, M. A formal definition of big data based on its essential features. *Library Review*, Emerald, v. 65, n. 3, p. 122–135, 04 2016. ISSN 0024-2535. Citado na página 12.
- MEDHAT, W.; HASSAN, A.; KORASHY, H. Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, Elsevier, p. 1093–1113, 2014. Citado na página 20.

MILLER, G. A. Wordnet: a lexical database for english. *Communications of the ACM*, p. 39–41, 1995. Citado na página 16.

MITCHELL, A.; GOTTFRIED, J.; MATSA, K. E. *Millennials and Political News*. 2015. Disponível em: <<http://www.journalism.org/2015/06/01/millennials-political-news/>>. Citado na página 11.

PADMAJA, S. Comparing and evaluating the sentiment on newspaper articles: A preliminary experiment. *Science and Information Conference*, IEEE, p. 789–792, 2014. Citado 2 vezes nas páginas 16 e 30.

PATIL, K.; FREDERIK, B. A measurement study of the content security policy on real-world applications. *International Journal of Network Security*, IEEE, p. 383–392, 2016. Citado na página 39.

PAUTASSO, C.; WILDE, E.; ALARCON, R. *REST: Advanced Research Topics and Practical Applications*. [S.I.]: Springer, 2014. Citado na página 17.

PESQUISA brasileira de mídia. Presidência da República. Secretaria de Comunicação Social, 2015. Citado na página 11.

POHL, K. *Requirements engineering: fundamentals, principles, and techniques*. [S.I.]: Springer Publishing Company, Incorporated, 2010. Citado na página 33.

PSOMAKELIS, E. et al. Comparing methods for twitter sentiment analysis. *CoRR*, abs/1505.02973, 2015. Disponível em: <<http://arxiv.org/abs/1505.02973>>. Citado na página 15.

SCHOLKOPF, B. et al. Comparing support vector machines with gaussian kernels to radial basis function classifiers. *IEEE Transactions on Signal Processing ( Volume: 45, Issue: 11, Nov 1997 )*, IEEE, p. 2758–2765, 1997. Citado na página 30.

Scrapy developers. *Architecture overview*. 2016. [Online and acessado em 10 de Agosto de 2016]. Disponível em: <<https://doc.scrapy.org/en/latest/topics/architecture.html>>. Citado 3 vezes nas páginas 6, 21 e 22.

STONEBRAKER, M. Sql databases v. nosql databases. *Commun ACM*, 2010. Citado na página 24.

TAN, A.-H. Text mining: The state of the art and the challenges. *Workshop on Knowledge Discovery from Advanced Databases*, 1999. Citado na página 18.

TICOM, A.; SOUZA, B. de; LIMA, L. de. Text mining and expert systems applied in labor laws. *Seventh International Conference on Intelligent Systems Design and Applications*, IEEE, 2007. Citado na página 12.

VINODHINI, G.; CHANDRASEKARAN, R. Sentiment analysis and opinion mining: a survey. *International Journal*, v. 2, n. 6, p. 282–292, 2012. Citado na página 32.

WANG, J.; GUO, Y. Scrapy-based crawling and user-behavior characteristics analysis on taobao. *Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*, IEEE, 2012. Citado na página 23.

WANJARI, Y. W. et al. Automatic news extraction system for indian online news papers. IEEE, 2014. Citado na página 12.

WIEBE, J. Tracking point of view in narrative. *Association for Computational Linguistics*, 1994. Citado na página 18.

YOONA, J. et al. Forensic investigation framework for the document store nosql dbms: MongoDB as a case study. *Digital investigation*, Elsevier, p. 53–65, 2016. Citado na página 24.