

INSTITUTO INFNET

NICOLAS DE SOUSA TEODOSIO E VICTOR HUGO NOVAIS
RODRIGUES

**ANÁLISE DE SENTIMENTO E MINERAÇÃO DE
OPINIÕES APLICADO NO TWITTER**

RIO DE JANEIRO

2016

INSTITUTO INFNET

NICOLAS DE SOUSA TEODOSIO E VICTOR HUGO NOVAIS
RODRIGUES

ANÁLISE DE SENTIMENTO E MINERAÇÃO DE OPINIÕES APLICADO NO TWITTER

Trabalho de Conclusão de Curso apresentado ao Programa de Graduação em Engenharia da Computação do Instituto Infnet como parte dos requisitos necessários à obtenção do título de Bacharel em Engenharia da Computação.

Orientador:
CASSIUS FIGUEIREDO

RIO DE JANEIRO

2016

NICOLAS DE SOUSA TEODOSIO E VICTOR HUGO NOVAIS
RODRIGUES

PySent: ANÁLISE DE SENTIMENTO E MINERAÇÃO DE DADOS APLICADO NO
TWITTER

Trabalho de Conclusão de Curso apresentado ao Programa de Graduação em Engenharia da Computação do Instituto Infnet como parte dos requisitos necessários à obtenção do título de Bacharel em Engenharia da Computação

Aprovada em XX agosto de 2016.

BANCA EXAMINADORA

Prof. Cassius Figueiredo, M.Sc. - Orientador
Instituto INFNET

A definir

A definir

Rio de Janeiro
2016

Às nossas famílias e amigos que durante esses cinco anos nos apoiaram nos momentos mais difíceis. Aos colegas e mestres que nos acompanharam nesta empreitada que no início eram desconhecidos, e hoje amigos que levaremos para o resto da vida.

Agradecimentos

Agradecemos inicialmente ao professor Cassius Figueiredo pelo total suporte nesse trabalho e por seu papel fundamental na nossa trajetória como alunos de Engenharia da Computação. Aos nosso pais, Rosângela de Sousa Teodosio, José Carlos Teodosio da Silva, Deise Luci Gouveia Novais e Paulo Roberto Rodrigues que nos proporcionaram uma base com valores familiares que nos permitiram seguir sempre nossos estudos, independentemente das dificuldades que surgiram pelo caminho.

Aos colegas de classe que fizemos durante estes 5 anos. Esperamos que, de uma forma ou de outra, todos alcancem o que desejam se seguirem o caminho do trabalho duro e da perseverança.

Aos amigos que nos ajudaram profissionalmente e academicamente, servindo de exemplo e dando conselhos que nos tornaram os profissionais que hoje somos: Ezequiel Bertti, João Leite, Thulio Costa, Felipe Salvini, Dalton Matos, Alan Dieguez, Jesue Souza e Cesar Frias.

Eu, Nicolas, agradeço a minha esposa Joyce Kelly Dias da Costa e minha irmã Jéssica de Sousa Teodosio por estar ao meu lado durante todo o curso e suportar todos os percalços que enfrentei até aqui.

Eu, Victor Hugo Novais Rodrigues, agradeço ao meu avô e grande exemplo de força, Manuel Ferreira Novais. Aos padrinhos Mauro Sérgio de Almeida Rubianes, Maysa Alexandrino dos Santos Rubianes e Maria Dionísia Pimentel dos Santos e aos tios Manoel de Carvalho Almeida e Vania Luci Gouveia Novais Almeida por todo carinho e apoio na minha criação e educação. A minha namorada Thays Scapin Coca pelo companheirismo e amor que me permitiram ir mais longe do que eu jamais sonhei.

Um agradecimento especial aos professores que cruzaram nossos caminhos durante o curso sempre transmitindo o conhecimento, experiência e sabedoria da maneira mais nobre possível.

Pedimos desculpas aqueles que por ventura tenham sido esquecidos. Esta graduação foi fruto do esforço de muitos e não só das duas pessoas que o encabeçam.

Resumo

Atualmente a internet e micro blogs em geral têm se tornado uma ferramenta de comunicação poderosa entre seus usuários. Bilhões de pessoas compartilham informações e opiniões a todo momento, fazendo desse espaço um ótimo campo de pesquisas comerciais, acadêmicas e sociológicas. Como esse fenômeno é relativamente recente, as pesquisas destinadas ao tema são consideradas novas, o Twitter, por exemplo foi criado apenas em 2006 e hoje possui 310 milhões de usuários ativos por mês .

Os principais desafios para aplicação da análise de sentimento estão relacionados a linguagens naturais sensíveis ao contexto, que não trazem resultados satisfatórios quando utilizam-se modelos matemáticos muito simples, sendo necessário um grande investimento de tempo em aperfeiçoar os modelos matemáticos disponíveis e adaptá-los à solução em questão.

O objetivo deste trabalho é explorar o potencial existente em pesquisas de opinião, que podem ser feitas através de análises nas comunicações utilizando a língua portuguesa nas redes sociais todos os dias, mostrando que o conteúdo escrito pelos usuários tem uma importância tão grande quanto a de números exibidos em redes sociais. A prática desse trabalho demandou uso de algoritmos matemáticos mais elaborados, aquisição de bases de palavras para o uso desse algoritmo, aplicação de um processo produtor-consumidor e acesso ao Twitter para coleta de dados. Foi realizada uma análise de caso sobre o Oscar 2016 para validar a efetividade desse trabalho.

O algoritmo utilizado foi o *Naive Bayes*, um algoritmo probabilístico, que realiza a análise de sentimento do oscar no momento de picos do evento, um deles sendo a premiação de melhor ator, atriz e filme. Nesses picos foram feitas análises de quantidade e taxa de polaridade. Seus resultados atestam a possibilidade da língua portuguesa como objeto de análise para a mineração de dados de forma relevante.

Palavras-chave: Análise de sentimento, mídias sociais, twitter, mineração de opiniões, processamento de linguagem natural, linguagens sensíveis a contexto, naive bayes.

Abstract

Today the internet and micro blogs in general have become a powerful communication tool among its users. Billions of people share information and opinions at all times, making this space a great field of commercial, academic and sociological research. As this phenomenon is relatively recent, research aimed at the topic is considered new, Twitter, for example was created only in 2006 and today has 310 million active users per month.

The main challenges for the application of the feeling analysis are related to context-sensitive natural languages, which do not yield satisfactory results when very simple mathematical models are used, requiring a great investment of time in improving the mathematical models available and adapting them to the solution in question.

The objective of this work is to explore the potential of opinion polls, which can be done through analysis in the communications using the Portuguese language in social networks every day, showing that the content written by users is as important as that of numbers displayed on social networks. The practice of this work required the use of more elaborate mathematical algorithms, acquisition of word bases for the use of this algorithm, application of a producer-consumer process and access to Twitter for data collection. A case study on the Oscar 2016 was conducted to validate the effectiveness of this work.

The algorithm used was the Naive Bayes, a probabilistic algorithm that performs the sentiment analysis of oscar at the moment of the peaks of the event, one of them being the best actor, actress and film award. At these peaks were made analyzes of quantity and polarity rate. Their results attest to the possibility of the Portuguese language as an object of analysis for data mining in a relevant way.

Palavras-chave: Sentiment analisys, social midia, twitter, opnion mining, natural language process, Context-sensitive languages, naive bayes.

Lista de Figuras

2.1	Visualizando as sete áreas práticas de mineração de texto e algumas de suas tarefas. [?]	5
2.2	O celular e a internet foram as armas dos rebeldes na Primavera Árabe. Fonte: Desconhecida	10
2.3	Papel das APIs integrando dados e serviços em diferentes plataformas. Fonte: http://www.programmableweb.com/	11
2.4	APIs mais utilizadas do mundo Fonte: SmartFile	12
2.5	O participante A (máquina) e o participante B (humano) se comunicam por texto com o participante C (juiz). Fonte: Wikipédia	15
3.1	Exemplo de arquitetura produtor-consumidor. Fonte: Apache ActiveMQ(http://activemq.apache.org/clustering.html)	21
3.2	Base SentiLex-PT antes da normalização	33
3.3	Base SentiLex-PT normalizada, dividindo em dois arquivos distintos os termos positivos e negativos	34
3.4	Amostragem da base de termos <i>Oscar 2016</i> normalizada, dividindo em dois arquivos distintos os termos positivos e negativos	35
4.1	Quantidade de tweets separados por polaridade do teste 1. Fonte: Própria	40
4.2	Quantidade de tweets separados por polaridade do teste 2. Fonte: Própria	41
4.3	Quantidade de tweets separados por polaridade do teste 3. Fonte: Própria	42
4.4	Quantidade de tweets separados por polaridade do teste 4. Fonte: Própria	43
4.5	Gráfico de comparação dos testes	43
4.6	Quantidade de tweets por tempo e polaridade. Fonte: Própria	45
4.7	Quantidade de tweets positivos diminuído pelos negativos pelo tempo. Fonte: Própria	45

4.8	Gráfico de bolhas referente aos eventos ocorridos entre 00:50 e 01:20 de probabilidade positiva	46
4.9	Gráfico de bolhas referente aos eventos ocorridos entre 00:50 e 01:20 de probabilidade negativa	47
4.10	Gráfico de bolhas referente aos eventos ocorridos entre 01:50 e 02:20 de probabilidade positiva	47
4.11	Gráfico de bolhas referente aos eventos ocorridos entre 01:50 e 02:20 de probabilidade negativa	48

Lista de Tabelas

3.1	Comparação entre bancos SQL e NoSQL	27
4.1	Exemplo de stemização	39
4.2	1º teste	39
4.3	2º teste	40
4.4	3º teste	41
4.5	4º teste	42
4.6	Comparando testes	43
B.1	Comparando testes com tempo	81

Lista de Abreviaturas e Siglas

API	Application Program Interface	1
PNL	Processamento de Linguagem Natural	13
IA	Inteligência Artificial	14
REST	<i>Representational state transfer</i>	12
SOAP	<i>Simple Object Access Protocol</i>	12
JSON	<i>Javascript Object Notation</i>	19
HTTP	<i>Hipertext Transfer Protocol</i>	19
XML	<i>Extensible Markup Language</i>	12
FIFO	<i>First-In-First-Out</i>	20
NLTK	<i>Natural Language Toolkit</i>	29
SQL	<i>Structure Query Language</i>	26
NoSQL	<i>Non Structure Query Language</i>	26

RDBMS <i>Relational database management system</i>	27
---	----

Sumário

1	Introdução	1
1.1	Motivação e Objetivos	2
1.2	Metodologia	2
1.3	Organização do trabalho	3
2	Referencial Teórico	4
2.1	Mineração de opinião	4
2.1.1	Sentimento	5
2.1.2	Desafios	6
2.1.3	Etapas	6
2.1.3.1	Coleta de dados	6
2.1.3.2	Classificação	7
2.1.3.3	Análise dos resultados	7
2.1.4	Aplicações práticas	7
2.1.4.1	Pesquisa de opinião sobre um produto	7
2.1.4.2	Análise sobre pessoas públicas	8
2.1.4.3	Bolsa de valores	8
2.1.5	Fontes de dados	8
2.1.5.1	Mecanismos de busca	8
2.1.5.2	Redes sociais	9
2.2	Twitter	9
2.2.1	Primavera Árabe	9

2.2.2	Análises de redes sociais	10
2.3	API	11
2.3.1	REST	12
2.3.2	SOAP	13
2.4	Processamento de linguagem natural	13
2.4.1	Definição	13
2.4.2	Teste de Turing	14
2.5	Classificador Naive Bayes	15
2.5.1	O Teorema de Bayes	16
2.5.2	Aplicação no trabalho	16
3	Proposta	18
3.1	Coleta de dados	18
3.1.1	Autenticação	19
3.1.2	Limite de requisições	19
3.1.3	Arquitetura	20
3.1.3.1	Produtor-consumidor	20
3.1.4	Busca	21
3.1.4.1	Parâmetros adicionais	21
3.1.4.2	O problema com a detecção automática de idioma do Twitter	22
3.1.4.3	Escalando de forma horizontal	22
3.2	Armazenamento	23
3.2.1	Banco de dados orientado a documento	26
3.2.2	Opções disponíveis	28
3.2.3	Armazenando <i>tweets</i>	28
3.3	Classificação	28
3.3.1	<i>Natural Language Toolkit</i> (NLTK)	29

3.3.2	<i>Stemming</i>	29
3.3.3	<i>Stopwords</i>	31
3.3.4	Construção da massa treino	32
3.3.5	Classificação	35
4	Resultados e análises	38
4.1	Cenários e parâmetros de teste	38
4.2	Técnicas	38
4.2.1	<i>Stemming</i>	39
4.2.2	<i>Stopwords</i>	39
4.3	Desenvolvimento do modelo de análise	39
4.4	Resultados	44
5	Conclusão	49
5.1	Trabalhos Futuros	50
	Apêndice A - Bases de texto utilizadas	52
A.1	<i>PUC</i>	52
A.1.1	Termos negativos	52
A.1.2	Termos positivos	53
A.2	<i>ReLi</i>	54
A.2.1	Termos negativos	54
A.2.2	Termos positivos	54
A.3	<i>SentiLex</i>	56
A.3.1	Termos negativos	56
A.3.2	Termos positivos	73
	Apêndice B - Lista de eventos do oscar em ordem cronológica e testes com tempo	80

Capítulo 1

Introdução

Através do fenômeno da popularização da Internet vivemos hoje um período conhecido como "Era da conhecimento"[?]. Nesse contexto, redes sociais mais conhecidas - como Facebook e Twitter - se tornaram bastante populares por permitirem a seus usuários acesso à um ambiente onde todos possuem voz e vez para se expressar e por consequência, para se informar sobre tudo que acontece no mundo. Através de Application Program Interface (API) disponibilizadas por essas redes sociais, possuímos fácil acesso a um grande volume de informação produzida pelos usuários e que podem ser utilizadas em pesquisas de opinião sobre um tema ou assunto específico. Tal cenário apresenta-se como uma grande oportunidade de pesquisa em áreas acadêmicas, sociais e comerciais.

Em Gomes [?] a mineração de texto é aplicada em busca de notícias sobre economia em Portugal. O trabalho concentra-se em monitorar sites relevantes que abordam notícias sobre a economia do país para representar o sentimento expresso no texto, através dos títulos das reportagens.

Em Pak e Paroubek [?] o Twitter é utilizado como fonte dados para análises de sentimento. O idioma de estudo escolhido foi o inglês, mas grande parte das técnicas apresentadas podem ser aplicadas em outras línguas.

Porém, quando o objeto de estudo é a língua portuguesa, nota-se que a mesma carece de trabalhos e implementações na área de mineração de opiniões e análise de sentimento. Alguns motivos explicam essa carência: poucos investimentos na área de ciência e engenharia da computação em nosso país e a grande dificuldade que a língua portuguesa apresenta ao ser interpretada através de processamento de linguagem natural. [?]

Alguns trabalhos utilizam o português como objeto de estudo, como por exemplo Tortella e Coelho [?]. Outros, se propõem a estudar um evento ou acontecimento, como

por exemplo as eleições presidenciais brasileiras no ano de 2010 [?], os protestos populares contra a corrupção ocorridos em 2013 [?] e a Copa do Mundo da FIFA Brasil em 2014 [?]. Nesses casos, *tweets* postados por usuários contendo *hashtags* referentes ao evento a ser estudado são monitorados e salvos numa base de dados ao longo do evento. Após o fim do mesmo, os dados são classificados utilizando um algoritmo previamente treinado e os resultados são analisados a fim de determinar a relevância, impacto e opiniões geral de acordo com a opinião dos usuários.

Neste trabalho, será aplicado o processo de mineração de opinião e análise de sentimento de forma semelhante a Pak e Paroubek [?], porém utilizando português como idioma de estudo. O objetivo é demonstrar como a mineração de opinião e análise de sentimento podem ser abordadas de forma abrangente, com a aplicação de técnicas generalistas. Além disso, é proposto um estudo de caso sobre a cerimônia do Oscar no ano de 2016 para também mostrar como um evento ou acontecimento pode ser estudado, aplicando técnicas específicas que tem como objetivo tornar a classificação mais especializada e precisa para este contexto, similar ao que foi feito em [?] [?] [?].

1.1 Motivação e Objetivos

Explorar o potencial contido no conteúdo digital gerado todos os dias em redes sociais por usuários de língua portuguesa é um grande desafio e serviu como motivação para este trabalho. Tais dados possuem informações valiosas de inúmeras maneiras - ainda inexploradas tanto comercialmente quanto academicamente na América Latina. O objetivo, portanto, é desenvolver uma prova de conceito utilizando um grande evento como estudo de caso para mineração de opinião, criando um *framework* de trabalho com foco em técnicas úteis no estudo da mineração de opinião em língua portuguesa, nas mais diversas aplicações existentes para o tema

1.2 Metodologia

Na pesquisa e análise dos dados apoia-se no método científico para alcançar conclusões. Para começar, foi necessário conhecer o tema - mineração de opinião - e familiarizar-se com o novo vocabulário. O estado da arte e artigos introdutórios ao tema serviram de alicerce.

De posse deste conhecimento uma análise do problema a ser resolvido abriu caminho

para o projeto, que é apresentado aqui e foi testado utilizando o evento de premiação dos Oscars 2016 e sua repercussão nas mídias sociais brasileiras como estudo de caso. As decisões de arquitetura e implementação são apresentadas através de comparações entre tecnologias e o seu impacto no produto final.

Por fim, foram coletados dados da classificação dos *tweets* de usuário de língua portuguesa durante o evento. Uma análise buscou identificar a eficácia da classificação e do processo de desenvolvimento aplicado, visando utilizá-lo como *framework* de trabalho para estudos posteriores.

1.3 Organização do trabalho

Este trabalho está estruturado em 5 capítulos da seguinte forma: no Capítulo 2, para embasamento teórico, são apresentados os conceitos de Mineração de opinião, Processamento de linguagem natural e classificação segundo Naive Bayes.

Em seguida, no Capítulo 3, são descritos os processos de coleta de dados, armazenamento e classificação, assim como suas respectivas arquiteturas, desafios e soluções. Em seguida, o Capítulo 4 apresenta os principais resultados obtidos com a análise feita em cima do estudo de caso do eventos Oscars 2016, assim como a análise dos valores encontrados.

Por fim, o Capítulo 5 conclui este trabalho, ressaltando os objetivos alcançados com as propostas. As principais vantagens e desvantagens da proposta são discutidas, assim como alguns trabalhos futuros que podem ser desenvolvidos.

Capítulo 2

Referencial Teórico

2.1 Mineração de opinião

É de conhecimento comum que há um acúmulo de dados por toda a internet. Artigos, informações de usuários, comportamento de usuários, são alguns tipos de informação que podem ser encontrados hoje na internet. Esse grande acúmulo não garante informações confiáveis ou uma análise correta sobre os dados, por isso há uma grande urgência para novas teorias computacionais e ferramentas que ajudem a analisar essa quantidade de dados gerados [?]. E dentro dessa enorme gama de dados, existem as informações adicionadas por usuários através de texto que remetem a suas reações a determinadas situações ou objetos.

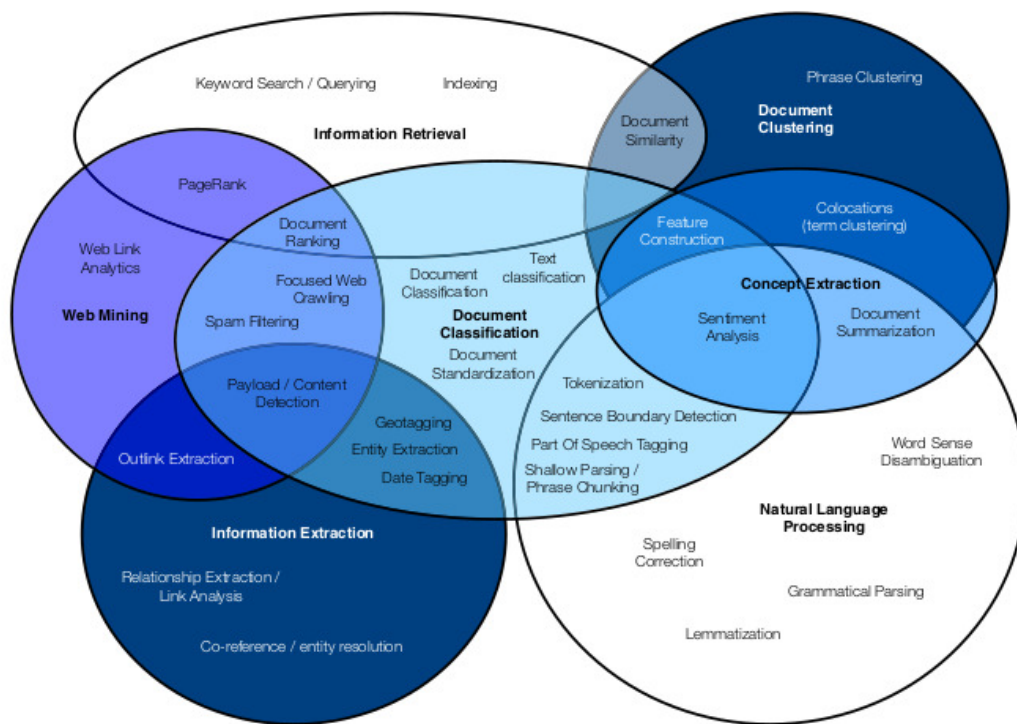


Figura 2.1: Visualizando as sete áreas práticas de mineração de texto e algumas de suas tarefas. [?]

De acordo com a figura 2.1, esse trabalho atua em uma área de interseção entre o processamento de linguagem natural, classificação de documento e extração de conceito.

2.1.1 Sentimento

De acordo com psicólogo Klaus R. Scherer, sentimento é um breve episódio da resposta sincronizada de todos os ou grande parte dos subsistemas orgânicos em resposta a um evento interno ou externo de grande significância[?]. Algumas outras definições utilizadas são:

- Ato ou efeito de sentir;
- Aptidão para receber as impressões;
- Sensação, sensibilidade;
- Consciência íntima;
- Faculdade de compreender, intuição e percepção;

A mineração de opinião, também conhecida como mineração de sentimento, análise de sentimento ou extração de opinião, é um campo dentro da mineração de dados [?] que tem como objetivo extrair o sentimento do texto escrito por uma pessoa, sem a interferência humana durante o processo.

2.1.2 Desafios

No campo de mineração de opinião, existem uma série de desafios que devem ser tidos como grandes pontos de atenção para quem deseja aplicar essa técnica de forma correta.

- Em blogs e redes sociais é comum encontrar textos com erros de ortografia ou escritos de forma informal, contendo gírias e abreviações comuns dentro da comunicação virtual;
- Dificuldade em discernir uma opinião ou um fato, especialmente quando existem opiniões embutidas em fatos;
- Os textos podem conter ironias e sarcarmos, que são especialmente difíceis de serem identificados e podem impactar os resultados;
- Um texto pode se referir à dois temas diferentes - política e ideologia, por exemplo - com opiniões diferentes sobre os mesmos, o que pode confundir a classificação;

2.1.3 Etapas

O processo de mineração de opinião consiste em 3 etapas: [?]

- Coleta de dados;
- Classificação;
- Análise dos resultados;

2.1.3.1 Coleta de dados

Nesta etapa é conduzida uma busca por opiniões nas mais diversas fontes que podem ser úteis: artigos, sites, comentários, anúncios dentre outras. Como explicado anteriormente, deve-se visar identificar se a informação coletada é uma opinião ou fato. Fatos podem

ser descartados imediatamente, porém opiniões apresentadas através de fatos, podem ser úteis.

Existem diversas maneiras de coletar sistematicamente fontes para extrair e armazenar os dados que serão utilizados, dentre elas as mais famosas estão o desenvolvimento *crawlers* - uma rotina sistemática capaz de varrer sites em busca de informações - e a utilização de APIs.

2.1.3.2 Classificação

A classificação é a alma do processo de mineração de opinião. Nesta etapa é determinada a polaridade do objeto de estudo em positivo, negativo e neutro.

Essa etapa é a principal responsável pela acurácia da análise. Por ser a etapa mais delicada do processo é onde ocorrem a maior parte dos erros. Existem diversas técnicas e ferramentas que ajudam a mitigar tais problemas que serão abordadas mais adiante, no Capítulo 3.

2.1.3.3 Análise dos resultados

A análise dos resultados envolve cruzar as informações de polaridade obtidas através texto com qualquer outra informação que exista sobre quem produziu aquela opinião. Desta forma, é possível, por exemplo, determinar qual gênero - masculino ou feminino - tem uma maior aceitação à um produto ou personalidade. As possibilidades para cruzar os dados e obter *insights* será proporcional a quantidade de informações coletadas durante o processo.

2.1.4 Aplicações práticas

Um algoritmo capaz de extrair opiniões de um texto pode ser aplicado em diversos cenários:

2.1.4.1 Pesquisa de opinião sobre um produto

Mineração de opinião pode ser usada por uma empresa para determinar se um certo produto lançado ao mercado atingiu a aceitação prevista, como forma de entender a percepção do público e guiar estrategicamente ações de marketing e relações públicas. Ainda é possível prospectar o sentimento associado a um produto antes mesmo do seu lançamento,

visando antecipar *insights* que podem ser valiosos durante o seu desenvolvimento.

2.1.4.2 Análise sobre pessoas públicas

Da mesma forma, é possível utilizar a mesma técnica e direcionar as análises para uma personalidade pública. Por exemplo, é possível determinar a aceitação ou rejeição de um político durante o mandato ou período de eleições, gerando dados que podem ser decisivos na definição de suas estratégias de campanha.

2.1.4.3 Bolsa de valores

Os números do mercado financeiro são uma consequência direta do sentimento que pessoas(investidores) possuem sobre uma empresa [?]. A opinião extraída de especialistas e sites de notícias podem ser usados como um dos fatores decisivos para compra e venda de ações.

2.1.5 Fontes de dados

É notório que estamos rodeados de dados dentro da Internet, porém dentro do campo de minerações de opiniões, existem algumas fontes que se destacam pela abrangência e diversidade dos dados.

2.1.5.1 Mecanismos de busca

É possível utilizar mecanismos de busca para obter opiniões sobre praticamente qualquer temática. Este método possui uma particularidade: mecanismos de busca como Google e Bing destacam certas páginas de acordo com motivos desconhecidos, o que pode influenciar os resultados obtidos. De forma geral, essa análise é apenas um reflexo do que está sendo buscado naquele momento.

Um exemplo da utilização de mecanismos de busca para mineração de opinião é o site whatdoesinternetthink.net [?], que utiliza como base os mecanismos de busca Google e Bing para determinar a opinião sobre um tema específico ou comparar dois temas entre si.

2.1.5.2 Redes sociais

O intenso compartilhamento de informações e opiniões que vemos hoje nas redes sociais serve como uma excelente fonte de dados para a mineração de opiniões por dois motivos: diversidade e abundância. Somando-se os usuários de Facebook e Twitter por exemplo, obtemos uma amostra considerável da população mundial à disposição para pesquisas.

Para este trabalho, o Twitter foi escolhido como base para a coleta de dados, por ser uma rede social focada em opiniões de usuários e pela grande facilidade que existe em consumir os seus dados através da API pública disponibilizada pelo mesmo.

2.2 Twitter

Contando com uma base ativa de usuários que ultrapassa 300 milhões [?], o Twitter é conhecido como um *microblog* fundado em março de 2006 por Jack Dorsey, Evan Williams e Biz Stone. Os usuários trocam mensagens de até 140 caracteres [?] em um ambiente de rede social, que tem como objetivo dar à todos o poder de compartilhar ideias e informações instantaneamente [?]. Após 10 anos de mercado, a empresa acumula números impressionantes: 300 bilhões de mensagens já foram compartilhadas por seus usuários, que em média enviam 500 milhões de *tweets* [?] - nome pelo qual as mensagens compartilhadas no microblog ficaram conhecidas na Internet - por dia.

Dentro do Twitter, O usuário pode fazer uso de *hashtags* - marcadores conhecidos do público de rede social, que servem como uma indexação para um tópico específico [?]. Apesar de simples, as *hashtags* pode ser usadas das mais diversas maneiras:

- Agrupar comentários e pensamentos acerca de um tema
- Estabelecer uma conexão entre dois tópicos
- Aproximar o usuários de um conteúdo relevante com auxílio de uma busca

2.2.1 Primavera Árabe

Um dos exemplos mais recentes e impressionantes de como as redes sociais desempenharam o papel de aproximar ideologias semelhantes e encorajar debates sociais profundos foi a Primavera Árabe - onda de manifestações e protestos que tiveram início em dezembro de 2010, tendo como cenário o Norte da África e Oriente Médio. Os principais alvos foram os

regimes ditatoriais e patriarcais que há muito tempo estavam no poder [?]. Redes sociais foram amplamente utilizadas para marcar encontros, debates e manifestações, além de mostrar para o mundo o que acontecia em tempo real, através do Twitter e outras redes sociais, como o YouTube.



Figura 2.2: O celular e a internet foram as armas dos rebeldes na Primavera Árabe. Fonte: Desconhecida

2.2.2 Análises de redes sociais

A análise de redes sociais ganhou incrível relevância nos campos de pesquisa social e comportamental[?] com o aumento de compartilhamento de informações dentro destes ambientes. Ao invés de analisar comportamentos individuais, atitudes e crenças, a análise de redes sociais foca sua atenção em entidades sociais ou atores interagindo entre si e como essas interações constituem uma estrutura que pode ser estudada e analisada.

Outro ponto levantado recorrentemente quando o assunto é análise de redes sociais é como ela pode ser útil para estudos de ordem micro ou macro. No nível *micro*, as análises destinam-se a examinar díades, tríades ou outros pequenos sub-grupos - conjuntos de duas, três ou mais atores sociais. No nível *macro*, o objeto de estudo são grandes redes de atores sociais. Todos os dados obtidos durante a coleta permitem segmentar os atores sociais de diversas formas - gênero, idade, religião, posição demográfica, entre outros. Por exemplo, os dados extraídos a partir da API do Twitter, tema abordada no Capítulo 3, nos permite entender como um usuário específico reagiu a uma *hashtag*. Da mesma forma, podemos olhar um cenário mais amplo, como por exemplo, todos usuários de uma região do país. As possibilidades de análise crescem e se tornam mais ricas conforme obtemos

mais informações sobre os atores no momento de suas interações sociais.

2.3 API

Uma API é um conjunto de rotinas estabelecidos por um software para a utilização de suas funcionalidades e acessos aos seus dados por outro software que não pretende fazer uso de sua implementação, apenas de seus serviços. Através dessa interface, capaz de fazer uma abstração dos dados e funcionalidades de um software, conectar-se a estes serviços se torna muito mais simples.

Outro ponto que demonstra a importância das APIs durante o desenvolvimento de software é a interoperabilidade. Atualmente, temos o mesmo serviço sendo oferecido em diferentes plataformas, como por exemplo *web*, *desktop* e *mobile*. Cada plataforma possui características e implementações diferentes, porém é possível que todas as plataformas utilizem as APIs como meio único de acesso a dados e serviços, promovendo uma padronização de protocolos e funcionalidades e serviços, além de alta reusabilidade de código.



Figura 2.3: Papel das APIs integrando dados e serviços em diferentes plataformas. Fonte: <http://www.programmableweb.com/>

O Twitter, nossa fonte de dados durante este trabalho, possui uma API pública que pode ser utilizada por qualquer usuário da rede social [?].

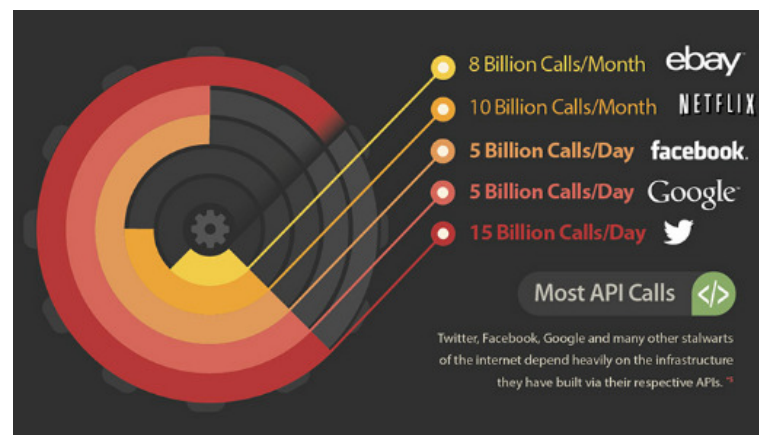


Figura 2.4: APIs mais utilizadas do mundo Fonte: SmartFile

Para efetuar uma comunicação eficiente com quem acessa à API, é necessário implementar um protocolo de acesso aos dados. Os protocolos *Representational state transfer* (REST) e *Simple Object Access Protocol* (SOAP) são os mais utilizados.

2.3.1 REST

O protocolo REST foi criado em 2000 por Roy Fielding [?] como parte de sua dissertação de doutorado na *University of California Irvine*. Suas principais vantagens são:

- Por ter sido criado dentro de um ambiente acadêmico, o objetivo do protocolo abraça a filosofia *open source* - que preza por projetos onde o código é aberto a todos para manutenção e colaboração [?];
- Fácil implementação e manutenção;
- Separa claramente a implementação do cliente e do servidor;
- A comunicação não é controlada por uma entidade única;
- A informação pode ser armazenada pelo cliente prevenindo múltiplas chamadas;
- Pode retornar a informação em múltiplos formatos - JSON, *Extensible Markup Language* (XML) , entre outros.

Por outro lado, o protocolo REST possui algumas limitações. Entre elas, podemos destacar:

- Só funciona em cima do protocolo HTTP;

- Autorização e recursos de segurança devem ser implementados à parte.

Baseado nessas características, o protocolo REST é comumente utilizado para APIs de aplicações *Web* e *Mobile*, como por exemplo, as APIs do Twitter, LinkedIn e Slack.

2.3.2 SOAP

Criado em 1998 por Dave Winer et al com colaboração da Microsoft, o protocolo SOAP concentra-se em endereçar necessidades do mercado corporativo. Como vantagem, o protocolo apresenta os seguintes aspectos:

- Segue uma abordagem mais formal, corporativa;
- Trabalha em cima de qualquer protocolo de comunicação, até mesmo assíncrono;
- Recursos de autorização e segurança incorporados de forma nativa;
- Pode ser descrito utilizando WSDL;

Entre suas principais desvantagens, podemos listar:

- Gasta-se muita banda tráfegando metadados
- Difícil implementação
- Pouco popular entre desenvolvedores *Web* e *Mobile*
- Retorna informação apenas em XML

Geralmente, o protocolo SOAP é mais utilizado em serviços financeiros, *gateways* de pagamento e serviços de telecomunicações.

2.4 Processamento de linguagem natural

2.4.1 Definição

Processamento de Linguagem Natural (PNL) baseia-se em modelos computacionais capazes de executar tarefas envolvem processar informações expresas em língua natural, como por exemplo, interpretação e tradução de textos. [?].

A pesquisa na área está voltada a quatro aspectos da comunicação essenciais:

- fonologia: estudo dos sons;
- morfologia: estudo da estrutura das palavras;
- semântica: estudo do significado;
- pragmática: estudo do significado aplicado a um contexto;

Neste trabalho, o PNL será aplicado à área da semântica e pragmática, responsável por estudar os elementos usados durante uma comunicação para se expressar através da língua (semântica) e a diversidade que pode surgir a partir de um contexto (pragmática). É também um estudo sobre como usuários de uma língua adquirem conhecimento sobre a mesma, através da comunicação oral ou escrita e como essa língua se altera ao longo do tempo.

Um dos grandes desafios da área é modelar o processamento de uma máquina para compreender uma estrutura tão complexa como uma linguagem. Existe um teste famoso na área de computação, o Teste de Turing, que levanta a questão "As máquinas podem pensar?". O teste fundamenta conceitos chave sobre a Inteligência Artificial (IA) , que serve como base para o PNL.

2.4.2 Teste de Turing

Introduzido pelo matemático britânico Alan Turing em seu artigo de 1950 "*Computing Machinery and Intelligence* [?]", o Teste de Turing explora a capacidade de um computador demonstrar comportamento inteligente equivalente ou indistinguível dos seres humanos.

O teste é composto por três elementos: dois seres humanos, sendo um participante e um juiz e um computador.

O juiz conversa em linguagem natural com um outro ser humano e uma máquina através de um canal de texto, composto por um teclado e uma tela que apresenta a conversa. Todos os participantes estão em ambientes separados. O juiz deve ser capaz de distinguir a máquina do ser humano, caso contrário, a máquina é considerada bem sucedida no teste. O objetivo não é analisar se a máquina é capaz de responder corretamente e sim dizer quão próximas as respostas da máquina foram das do ser humano.

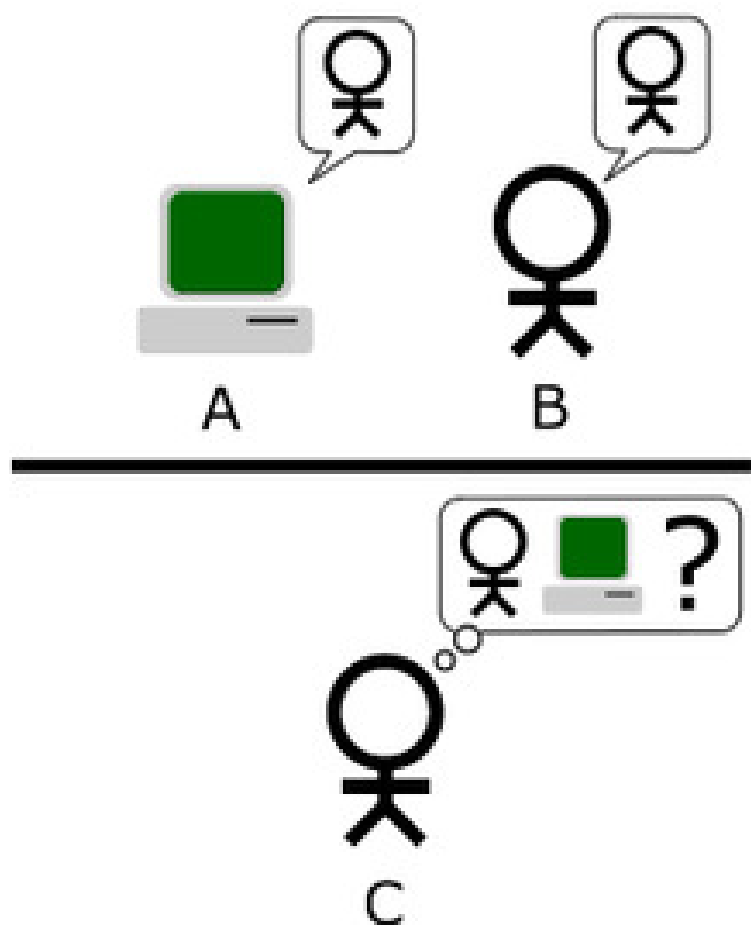


Figura 2.5: O participante A (máquina) e o participante B (humano) se comunicam por texto com o participante C (juiz). Fonte: Wikipédia

2.5 Classificador Naive Bayes

O classificador conhecido como *Naive Bayes* é um algoritmo probabilístico baseado no Teorema de Bayes que não considera que eventuais dependências possam existir. Por este motivo, suas suposições são nomeadas "ingênuas" - de onde surge o nome *naive* - o que lhe confere uma maior simplicidade e um desempenho maior, em relação a outros algoritmos de classificação [?]. É um método popular para categorização de textos, como por exemplo a classificação de *e-mails* em legítimos ou *spam* - e-mails inúteis que são enviados na esperança que o receptor compre algum produto ou serviço. [?]

2.5.1 O Teorema de Bayes

O Teorema de Bayes permite inferir qual é a probabilidade de um evento A dado um evento B e pode ser expressado pela seguinte equação:

$$P(A | B) = \frac{P(B | A) P(A)}{P(B)},$$

onde A e B são eventos.

- $P(A)$ e $P(B)$ são probabilidades de A e B sem considerar a relação entre ambos;
- $P(A|B)$, uma probabilidade condicional, é a probabilidade de observar o evento A, dado que o evento B ocorreu.
- $P(B|A)$ é a probabilidade de observar o evento B, dado que o evento A ocorreu.

Suponha que queremos saber a probabilidade de um indivíduo possuir câncer, sem saber nada sobre o indivíduo. Porém, sabemos que a chance de um indivíduo estar infectado com tal câncer é de 1%, ou seja, $P(A)$. Em seguida, suponha que esta pessoa tenha 70 anos de idade e que essa probabilidade é de 0,2% e que 0,5% das pessoas doente possuem 70 anos de idade ou $P(B)$. Se assumirmos que a incidência de câncer e a idade estão relacionadas, podemos utilizar esta informação para melhor medir as chances desta pessoa estar doente. Logo, queremos saber a probabilidade de uma pessoa estar doente quando a mesma possui 70 anos de idade, ou $P(A|B)$.

$$\left(\frac{0,5}{100} \times \frac{1}{100}\right) \div \frac{0,2}{100} = \frac{2,5}{100}$$

Portanto, o resultado do teorema demonstra que possuir 70 anos de idade aumenta a chance de uma pessoa ser portadora de câncer, apesar desta probabilidade ainda ser baixa.

2.5.2 Aplicação no trabalho

Neste trabalho o *Naive Bayes* é utilizado para categorizar *tweets* extraídos do Twitter em positivos, neutros ou negativos. A diferença deste algoritmo para o Teorema de Bayes é assumir que a posição das palavras - eventos da probabilidade - que aparecem no texto não importa para determinar o resultado final.

Como visto em [?] o algoritmo calcula qual a probabilidade de uma frase, denominada documento pertencer a uma determinada classe(polaridade) $P(C/D)$, a partir da probabilidade de $P(C)$ do documento pertencer a esta classe e das probabilidades condicionais de cada termo t_k ocorrer em um documento da mesma classe. O algoritmo tem como objetivo encontrar a melhor classe para um documento maximizando a probabilidade conforme a equação abaixo, onde n_d é o número de termos no documento d .

$$C_{map} = \operatorname{argmax}_{c \in C} P(c|d) = \operatorname{argmax}_{c \in C} P(c) \prod_{k=1}^{n_d} P(t_k/d)$$

Capítulo 3

Proposta

Como visto no Capítulo 2, existe uma corrente dentro da Mineração de Opinião que vem desenvolvendo maneiras de explorar o conteúdo digital gerado pela nossa sociedade todos os dias em redes sociais, através de técnicas utilizando Processamento de Linguagem Natural e *Machine Learning*, principalmente. Com este fato surge a oportunidade de explorar novas ferramentas na solução de problemas que envolvem pesquisas de opinião de forma geral.

Neste trabalho propõem-se um *framework* que torna possível fazer pesquisas de opiniões em língua portuguesa sobre qualquer tema que seja rastreável a partir de uma *hashtag* no Twitter. Para tal é necessário que o *framework* criado seja capaz de:

1. Coletar *tweets* escritos em língua portuguesa que contenham uma determinada *hashtag*;
2. Armazenar as mensagens em uma base de dados;
3. Classificar as mensagens de acordo com a polaridade: negativo, neutro e positivo;
4. Extrair *insights* que auxiliem a tomada de decisão a partir da massa de dados classificada;

3.1 Coleta de dados

A plataforma do Twitter conecta aplicações e sites com seus dados através de diversos serviços. Para este trabalho, a principal fonte de dados será sua API REST, que possui uma excelente documentação disponível em [?]. Através dela é possível acessar informações de usuários e *tweets*, assim como escrever novas mensagens. Além disso, a API conta

com um mecanismo de busca poderoso, que será fundamental para a coleta de dados. Os dados são entregues no formato *Javascript Object Notation* (JSON).

3.1.1 Autenticação

Para que ter acesso à API antes é necessário possuir uma conta no Twitter e criar uma aplicação - através do próprio site [?] - que utilizará o protocolo de autenticação OAuth[?] para acessar os dados do Twitter se passando pelo usuário em questão. O objetivo do protocolo OAuth é permitir que uma aplicação se autentique em outra "em nome de um usuário". A aplicação pede permissão de acesso ao usuário, que possui a escolha de conceder permissão ou não. Um ponto importante: o usuário não precisa informar a sua senha para se autenticar, portanto a permissão continua vigente caso a senha do usuário se altere, o que permite que a aplicação não precise de manutenção neste caso, tornando-a mais resiliente. A autenticação por meio do OAuth necessita de três passos:

1. Aplicação cliente obtém chave de autenticação;
2. Usuário autoriza aplicação cliente na aplicação servidora;
3. Aplicação cliente troca a chave de autenticação pela chave de acesso;

Após o processo de criação da aplicação, é criado um *token* de acesso que deve ser utilizado pelo sistema que deseja se autenticar no Twitter em nome de um usuário. Este *token* deve ser incorporado em cada requisição à API do Twitter para autenticar a mesma e dizer ao Twitter qual é a fonte do acesso.

3.1.2 Limite de requisições

A fim de evitar grande concentração de requisições em seus serviços, o Twitter implementa um limitador em sua API [?]. São permitidas até 180 requisições por janela, que dura 15 minutos. Caso o limite seja ultrapassado, o serviço passa a retornar um erro na resposta, até que a "janela" de 15 minutos se renove.

A partir da versão 1.1 da API, novos cabeçalhos *Hipertext Transfer Protocol* (HTTP) são retornados provendo *feedback* sobre os limites para requisição. Este recurso permite que o código consiga entender em que momento da janela se encontra, quantos requisições ainda podem ser feitas neste período de tempo e quanto é necessário esperar para poder fazer novas requisições. Os cabeçalhos em questão são:

- *X-Rate-Limit-Limit*: A faixa limite para o requisição em questão;
- *X-Rate-Limit-Remaining*: O número de requisições que ainda restam para a janela de 15 minutos;
- *X-Rate-Limit-Reset*: O tempo restante dentro da janela de requisições atual, dado em segundos.

3.1.3 Arquitetura

Neste trabalho, como o objetivo é coletar *tweets* postados sobre uma *hashtag* em tempo real para utilizá-los como matéria-prima para análise de sentimento, é muito importante aproveitar ao máximo cada janela de requisições. Por este motivo, o sistema que coleta os dados da API do Twitter foi inspirado no modelo produtor-consumidor[?] visando minimizar as perdas que podem acontecer em momentos de pico - como o começo ou clímax do evento, onde o volume de mensagens é maior, como veremos a frente no Capítulo 4 - e se necessário, escalar de forma simples durante os mesmos.

3.1.3.1 Produtor-consumidor

O problema descreve dois processos, o produtor e o consumidor, que compartilham um recurso em comum usado como uma fila - um tipo particular de coleção de dados onde a primeira a entidade a entrar é a primeira a sair ou *First-In-First-Out* (FIFO) . A função do produtor é gerar trabalho a ser executado pelo consumidor. O volume de trabalho gerado e executado pelo sistema é controlado pela fila, que armazena as entidades ou "tarefas" a serem executadas. Essa abordagem permite que o sistema escale apenas até a sua capacidade, visto que a fila possui um tamanho fixo que caso seja ultrapassado, pode simplesmente descartar as mensagens adicionadas após este momento. Outra característica importante é a escalabilidade. Conforme os processos produtor e consumidor evoluem, surge a necessidade de aumentar a quantidade de produtores ou consumidores de forma independente.

Neste trabalho, para explorar o potencial máximo da janela de requisições foi criado um processo produtor que envia para a fila mensagens para que consumidor acesse à API do Twitter de forma que sejam feitos sempre as 180 requisições que são permitidas no intervalo de 15 minutos. Assim a responsabilidade de cada processo fica bem definida - o primeiro responde pelo volume de requisições e o segundo por realizar a requisição e

entender a resposta. Para definir qual intervalo de tempo deveria ser usado para que o produtor envie mensagens à fila, foi feita uma conta simples:

$$\frac{180}{15} = 12 \text{ requests}_{/minuto}$$

Logo, o produtor precisa adicionar uma mensagem na fila a cada 5 segundos, para que o limite de 180 requisições seja respeitado.

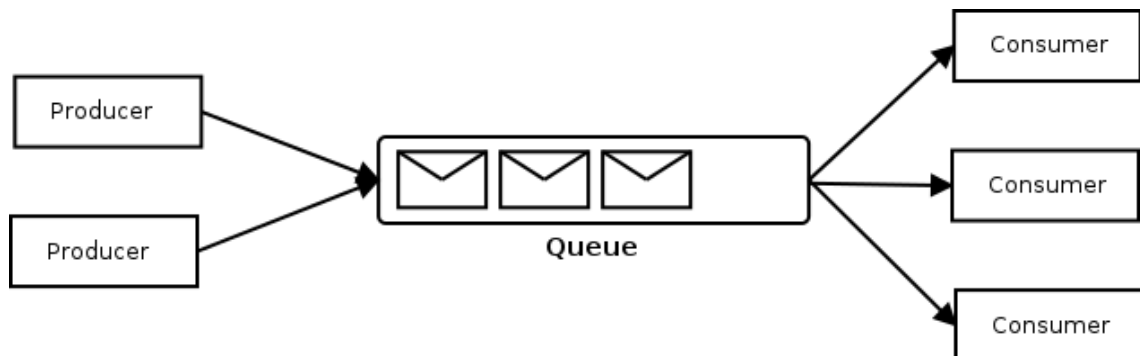


Figura 3.1: Exemplo de arquitetura produtor-consumidor. Fonte: Apache ActiveMQ(<http://activemq.apache.org/clustering.html>)

3.1.4 Busca

Dentre os principais serviços da API do Twitter está a busca. Com ela, é possível consultar de diversas formas os principais *tweets* ou mais recentes. Dentro de sua documentação, existe um guia completo de como utilizar a API para extrair os resultados desejados [?] das mais diversas formas.

3.1.4.1 Parâmetros adicionais

Como abordado acima, a API possui diversos parâmetros que podem ser usados para que o usuário chegue a um conjunto de dados mais próximo da sua necessidade:

- **result_type**: permite escolher se o resultado da busca será representado pelos *tweets* mais populares (*popular*) ou mais recentes (*recent*);
- **geocode**: permite buscar por uma determinada latitude, longitude e raio, respectivamente, separando-os por vírgula. ex: geocode=-22.912214,-43.230182,1km;
- **lang**: restringe os *tweets* buscados a um idioma específico. ex: lang=pt;

- *since_id* , *max_id* , *count* e *until*: possibilita iterar através dos resultados quando existe um grande número de *tweets* a percorrer. De acordo com a concorrência e o volume, esta tarefa pode ficar mais complicada. Uma leitura recomendada se encontra em [?].

Como o objetivo deste trabalho é coletar novos *tweets* conforme eles vão sendo postados, foi necessário utilizar apenas dois parâmetros da API de busca: *count* e *since_id*. O primeiro tem o objetivo de garantir o número máximo de registros retornados pela API e o segundo define de onde se pretende partir para buscar novos *tweets*, evitando que mensagens repetidas sejam coletadas.

```
https://api.twitter.com/1.1/search/tweets.json?q=#oscars2016&count=100&since_id=123456789
```

Neste caso, a API retornará os 100 *tweets* publicados desde o *tweet* com identificador (*id*) "123456789".

3.1.4.2 O problema com a detecção automática de idioma do Twitter

O escopo deste trabalho determina que o objeto de estudo são apenas mensagens escritas em língua portuguesa. Como forma de obter somente *tweets* escritos em língua portuguesa é preciso utilizar o parâmetro *lang*, por exemplo:

```
https://api.twitter.com/1.1/search/tweets.json?q=#oscars2016&lang=pt&result_type=recent
```

Porém, realizando alguns testes na API do Twitter, foi detectado que ao submeter algum *tweet*, alguma rotina dentro do próprio Twitter atribui um idioma à mensagem automaticamente. Nos testes conduzidos durante este trabalho foi identificado que em mensagens curtas - algo recorrente no Twitter - o algoritmo apresenta resultados aquém do esperado na identificação do idioma, visto que as poucas palavras contidas na mensagem podem ser comuns a mais de um idioma. Por conta disso foi decidido que o filtro de idioma não seria utilizado. Esta decisão pode mudar de acordo com o evento monitorado ou com o escopo do estudo.

3.1.4.3 Escalando de forma horizontal

Por uma questão de desenho da API de busca, cada requisição é capaz de trazer no máximo 100 *tweets*, o que nos dá ao total uma carga máxima de até 1200 novos *tweets*

por minuto. Para as análises feitas durante este trabalho este número se mostrou mais do que suficiente. Como dito anteriormente, se fosse necessário escalar este sistema para monitorar um evento maior, seria necessário apenas obter um saldo maior de requisições junto a API do Twitter - adicionando mais *tokens* de usuário e criando uma espécie de "rodízio" de autenticações, por exemplo - e escalar o número de consumidores do processo de acordo com a demanda. Uma boa maneira de detectar se isso seria necessário é acompanhar quantos *tweets* novos são coletados a cada requisição. Como utilizamos o *id* do último *tweet* capturado como referência para os novos, se a cada requisição o número de novos *tweets* com grande frequência coincidir com o limite da API, temos um indício de que o volume de novas mensagens no Twitter está excedendo a capacidade do sistema de coletá-las e que o excedente está sendo perdido.

3.2 Armazenamento

A resposta da API de Busca é dada no formato JSON e cada objeto - que corresponde a cada *tweet* - segue o seguinte formato e conta com diversas informações sobre o mesmo:

```
1 {
2   "_id" : "56d388096861353c1b061fe2",
3   "contributors" : null,
4   "truncated" : false,
5   "text" : "Leonardo DiCaprio com o #oscars eh igual a Katy Perry com o
6           Grammy #OscarsNaTNT",
7   "is_quote_status" : false,
8   "in_reply_to_status_id" : null,
9   "id" : 704091511902834688,
10  "favorite_count" : 0,
11  "source" : "<a href=\"http://twitter.com/download/android\" rel=\"
12             nofollow\">Twitter for Android</a>",
13  "created_at_datetime" : ISODate("2016-02-28T20:51:12.000Z"),
14  "retweeted" : false,
15  "coordinates" : null,
16  "created_at_timestamp" : 1456714272.0000000000000000,
17  "entities" : {
18    "symbols" : [],
19    "user_mentions" : [],
```

```
18     "hashtags" : [{
19         "indices" : [ 24, 31 ],
20         "text" : "oscars"
21     },
22     {
23         "indices" : [ 66, 78],
24         "text" : "OscarsNaTNT"
25     }],
26     "urls" : []
27 },
28 "in_reply_to_screen_name" : null,
29 "id_str" : "704091511902834688",
30 "retweet_count" : 0,
31 "in_reply_to_user_id" : null,
32 "favorited" : false,
33 "user" : {
34     "follow_request_sent" : false,
35     "has_extended_profile" : true,
36     "profile_use_background_image" : false,
37     "id" : 2786117482,
38     "verified" : false,
39     "profile_text_color" : "000000",
40     "profile_image_url_https" : "https://pbs.twimg.com/profile_images/
41         700450356141158404/xA-mRqp7_normal.jpg",
42     "profile_sidebar_fill_color" : "000000",
43     "is_translator" : false,
44     "entities" : {
45         "description" : {
46             "urls" : []
47         },
48     },
49     "followers_count" : 156,
50     "protected" : false,
51     "location" : "Um lugarzinho no fim do mundo",
52     "default_profile_image" : false,
```

```
52     "id_str" : "2786117482",
53     "lang" : "pt",
54     "utc_offset" : -28800,
55     "statuses_count" : 6171,
56     "description" : "Uuuuum ta estilosa!",
57     "friends_count" : 192,
58     "profile_background_image_url_https" : "https://abs.twimg.com/images/
        themes/theme1/bg.png",
59     "profile_link_color" : "9266CC",
60     "profile_image_url" : "http://pbs.twimg.com/profile_images/
        700450356141158404/xA-mRqp7_normal.jpg",
61     "notifications" : false,
62     "geo_enabled" : false,
63     "profile_background_color" : "000000",
64     "profile_banner_url" : "https://pbs.twimg.com/profile_banners/
        2786117482/1456444936",
65     "profile_background_image_url" : "http://abs.twimg.com/images/themes/
        theme1/bg.png",
66     "name" : "Padeira Estilosa",
67     "is_translation_enabled" : false,
68     "profile_background_tile" : false,
69     "favourites_count" : 6095,
70     "screen_name" : "naycordeir",
71     "url" : null,
72     "created_at" : "Fri Sep 26 18:43:47 +0000 2014",
73     "contributors_enabled" : false,
74     "time_zone" : "Pacific Time (US & Canada)",
75     "profile_sidebar_border_color" : "000000",
76     "default_profile" : false,
77     "following" : false,
78     "listed_count" : 1
79 },
80 "geo" : null,
81 "in_reply_to_user_id_str" : null,
82 "lang" : "pt",
```



```
83  "created_at" : "Sun Feb 28 23:51:12 +0000 2016",
84  "metadata" : {
85    "iso_language_code" : "pt",
86    "result_type" : "recent"
87  },
88  "in_reply_to_status_id_str" : null,
89  "place" : null
90 }
```

Dentro desta resposta, existem diversos dados que podem ser úteis para análises em cima dos *tweets* e armazená-los é extremamente valioso.

3.2.1 Banco de dados orientado a documento

Existem diversas soluções de banco de dados disponíveis no mercado. Nos últimos anos, uma delas se tornou especialmente popular[?]: os bancos de dados orientado a documento. Tais bancos de dados são uma das principais categorias de bancos conhecidos como *Non Structure Query Language* (NoSQL) que consiste em organizar os dados de forma "não-relacional", através de documentos, gráficos, chave-valores e colunas. Bancos NoSQL são conhecidos pela facilidade de modelagem e desenvolvimento, alto desempenho de leitura e escrita, alta disponibilidade e resiliência. Isso não significa que bancos *Structure Query Language* (SQL) são obsoletos ou piores, porém existem aplicações claras onde cada um desempenha um melhor papel. Podemos apontar algumas comparações, como por exemplo:

	Banco de dados relacional	Banco de dados NoSQL
Modelagem	O modelo relacional normaliza dados em estruturas tabulares conhecidas como tabelas, que consistem em linhas e colunas. Um esquema ou <i>scheme</i> define estritamente as tabelas, colunas, índices, relações entre tabelas e outros elementos do banco de dados.	Bancos de dados não relacionais (NoSQL) normalmente não aplicam um esquema. Geralmente, uma chave de partição é usada para recuperar valores, conjuntos de colunas ou documentos semi-estruturados JSON, XML ou outros que contenham atributos de itens relacionados.
Desempenho	O desempenho normalmente depende do subsistema do disco. A otimização de consultas, índices e estrutura de tabela é necessária para alcançar máximo desempenho.	Desempenho geralmente é uma função do tamanho do <i>cluster</i> do hardware subjacente, da latência de rede e da aplicação que faz a chamada.
Escala	Mais fácil de aumentar a escala "verticalmente" com hardware mais rápido. Outros investimentos são necessários para tabelas relacionais para abranger um sistema distribuído.	Projetado para aumentar a escala "horizontalmente" usando <i>clusters</i> distribuídos de hardware de baixo custo para aumentar a transferência sem aumentar a latência.
APIs	As solicitações para armazenar e recuperar dados são comunicadas usando consultas compatíveis com SQL. Essas consultas são analisadas e executadas por <i>Relational database management system</i> (RDBMS) .	APIs baseadas em objetos permitem que desenvolvedores de aplicações armazenem e restaurem facilmente estruturas de dados na memória. As chaves de partição permitem que os aplicativos procurem pares de chave-valor, conjuntos de colunas ou documentos semi-estruturados contendo objetos e atributos de aplicativos serializados.

Tabela 3.1: Comparação entre bancos SQL e NoSQL

3.2.2 Opções disponíveis

O movimento de adoção de bancos *NoSQL* está bastante enraizada no mundo *open source*. Alguns projetos como Voldemort[?], MongoDB[?], Tokyo Cabinet[?] e CouchDB[?]. Apesar de uma grande quantidade de opções *open source*, o movimento ganhou muita força com a publicação de duas publicações sobre implementações proprietárias: o Google BigTable[?] e o Amazon Dynamo[?]. Para este trabalho, a opção escolhida foi o MongoDB, altamente popular na comunidade *open source* e com bastante material disponível com melhores práticas de criação, manutenção e configuração.

3.2.3 Armazenando *tweets*

O objetivo deste trabalho é monitorar eventos através das *hashtags*. E para cada uma delas é criada uma coleção - nome dado a um conjunto de documentos - dentro do banco de dados. O nome da coleção é dado pela *hashtag* monitorada, como por exemplo #Oscars2016.

No momento onde o processo consumidor - responsável por fazer requisições na API e lidar com o retorno - é ligado ocorre a criação da coleção, caso a mesma não exista, e dentro dela começam a ser armazenados os *tweets* retornados pela API, obedecendo ao mesmo formato enviado pelo Twitter. Antes do armazenamento no banco alguns campos a mais são adicionados, mas vamos entrar neste detalhe apenas na seção sobre Classificação.

3.3 Classificação

- Aplicar técnicas de normalização no texto. As mesmas devem ser específicas para a língua portuguesa;
- Construir base de palavras e termos classificados utilizadas como insumo para o modelo matemático;
- Preparar uma massa de treino para validar o modelo matemático antes da execução;
- Calibragem do algoritmo
- Salvar informações sobre a classificação dentro do documento

O processo de classificação é a alma deste projeto. É ele que determina qual sentimento será possível extrair de um *tweet* através conteúdo altamente resumido do mesmo. Durante o estudo e aperfeiçoamento deste processo foram utilizadas algumas técnicas que visam:

- Diminuir a variabilidade de aplicações para uma palavra dentro da língua portuguesa. ex: aplicação de gênero, plurais e singulares, conjugações verbais, entre outras;
- Excluir do texto termos que são inúteis semanticamente, como por exemplo conjunções, preposições e marcações de usuários, conhecidas como menções ou *mentions*;
- Treinar o algoritmo de classificação com uma base de palavras pertencentes ao mesmo universo e temática do estudo, permitindo que o algoritmo previamente conheça o contexto e força que alguns termos vão exercer sobre o texto.

3.3.1 *Natural Language Toolkit (NLTK)*

Durante o estudo foram utilizadas algumas ferramentas que auxiliaram o processo de classificação. A principal delas foi a NLTK [?], uma biblioteca em Python construída como suíte de soluções para trabalhar com linguagem natural. O NLTK conta com mais de 50 corpus linguísticos - conjunto de registros orais de uma determinada língua que servem como base para análises - além de diversas ferramentas para classificação, marcação e análise de texto. Seguindo uma documentação específica para aplicação em língua portuguesa[?], foram utilizados os seguintes corpus do NLTK durante este trabalho:

- Floresta Sintá(c)tica Corpus versão 7.4: uma colaboração entre a Linguateca[?] e o projeto VISL[?]. Contém textos em português (do Brasil e de Portugal) analisados automaticamente pelo analisador sintático PALAVRAS (Bick 2000) e revistos por linguistas.
- Mac Morpho: Palavras extraídas a partir de textos do jornal Folha de São Paulo
- Machado: Obra completa do escritor Machado de Assis

3.3.2 *Stemming*

Dentro da língua portuguesa, uma mesma palavra pode variar de diversas formas de acordo com a sua aplicação. Para citar um exemplo: uma substantivo pode variar em gênero - masculino ou feminino - número - singular ou plural - e grau - aumentativo ou diminutivo. Seguindo a mesma vertente, verbos podem mudar de acordo com a conjugação dos tempos verbais, como por exemplo, aplicações no pretérito, presente ou futuro. Porém, toda palavra da língua portuguesa possui um "núcleo" que não muda independente do

gênero, número, grau ou conjugação. Essa raiz da palavra é conhecida como radical - que significa relativo ou pertencente à raiz ou à origem.

Para que a classificação seja eficiente em definir o sentimento contido nas mensagens extraídas a partir do Twitter é necessário que os termos existentes no mesmo sejam palavras já classificadas previamente em negativas ou positivas nas bases que foram escolhidas para treinar nosso algoritmo, como veremos mais a frente. Extrair os radicais tanto das palavras contidas no *tweet* quanto das bases classificadas que utilizamos é uma técnica fundamental, dado que com ela se remove grande parte das variações linguísticas de um termo sem alterar o seu significado. A adição desta técnica no fluxo de normalização de texto foi feita num estágio tardio do projeto e gerou um cenário de teste que é apresentado no Capítulo 4, onde pode-se ver uma diminuição considerável no número de "falso neutros", ou seja, ocorrências onde o algoritmo falhava em classificar mensagens por falta de dados e por consequência as classificava como neutras.

O NLTK possui duas implementação de *stemmers* em sua biblioteca: a primeira é o *Poter Stemmer* - algoritmo escrito por Martin Porter e publicado em 1980. O algoritmo se tornou extremamente popular e com isso se tornou o algoritmo padrão para processos de *stemming* em língua inglesa. Para abranger outras línguas, outros algoritmos foram escritos, porém muitos deles possuíam falhas que dificultaram sua adesão. Por conta disso, durante os anos 2000 Porter expandiu seu trabalho criando o Snowball[?], um *framework* para desenvolvimento de algoritmos de *stemming*, um dos quais será utilizado neste trabalho para realizar o processo em língua portuguesa.

```
1 In[1]: from nltk.stem.snowball import SnowballStemmer
2 In[2]: print(" ".join(SnowballStemmer.languages))
3
4 Out[1]: danish dutch english finnish french german hungarian italian
5 norwegian porter portuguese romanian russian spanish swedish
```

```
1 In[1]: from nltk.stem.snowball import SnowballStemmer
2 In[2]: Stemmer("portuguese", ignore_stopwords=True)
3 In[3]: print(stemmer.stem("mineracao de opiniao"))
4
5 Out[1]: u"miner opin"
```

3.3.3 *Stopwords*

A composição de um *tweet* por muitas vezes possui palavras que são semanticamente nulas ou até mesmo nocivas para o nosso algoritmo de classificação. Por conta disso, uma das primeiras técnicas aplicadas em um texto antes da classificação consiste em conduzir uma normalização nos elementos contidos em uma sentença.

Palavras que não possuem valor semântico para uma sentença são conhecidas como "palavras vazias" ou *stopwords*. O NLTK possui um vasto dicionário de palavras vazias graças aos corpus de língua portuguesa adicionados à ele anteriormente. Com base neste dicionários, é possível excluir palavras vazias da seguinte forma:

```
1 In[1]: text = "Veja a lista completa dos vencedores do Oscar 2016"
2 In[2]: [i for i in text.split() if i not in stopwords.words("portuguese")]
3
4 Out[1]: ["Veja", "lista", "completa", "vencedores", "Oscar", "2016"]
```

Além disso, foi necessário adicionar inteligência ao nosso normalizador para também excluir palavras vazias específicas para *tweets*, como por exemplo as *hashtags*, menções a outros usuários - as *mentions* - e *links* para outras páginas, vídeos ou imagens.

```
1 In[1]: TWITTER_STOPWORDS = ["@", "RT", "http"]
2 In[2]: def clean_stopwords(self, text):
3     # Cleaning portuguese stopwords
4     splitted = [i for i in text.split() if i not in stopwords.words("portuguese")]
5     cleaned_splitted = []
6
7     # Cleaning twitter stopwords
8     for word in splitted:
9         cleaned_splitted.append(word)
10
11     for twitter_stopword in TWITTER_STOPWORDS:
12         if word.startswith(twitter_stopword):
13             cleaned_splitted.remove(word)
14
15     return " ".join(cleaned_splitted)
16
17 In[3]: text = "@tntbr esta comecando agora a transmissao ao vivo do #Oscars2016
18         http://bit.ly/oscarstnt"
19 In[4]: clean_stopwords(text)
```

19

20

```
Out[1]: "esta comecando agora transmissao vivo"
```

3.3.4 Construção da massa treino

Durante o processo de pesquisa de bases de termos classificadas previamente em positivos e negativos, foram encontradas algumas fontes interessantes a se utilizar como bases de dados confiáveis para a mineração de opinião em língua portuguesa.

A primeira delas veio de um artigo escrito por Cláudia Freitas, estudante da Pontifícia Universidade Católica do Rio de Janeiro [?]. A temática do artigo gira em torno da construção de um léxico da afetividade para processamento computacional do português, que possui alta sinergia com a proposta deste trabalho. Esta base de dados foi apelidada durante o trabalho de "PUC". Outra base de dados utilizada durante este trabalho - apelidada de SentiLex-PT - foi encontrada durante a leitura de um artigo intitulado *Building a sentiment lexicon for social judgement mining* [?], escrito em 2012 por Mário J. Silva, pesquisador Sênior do Instituto de Engenharia de Sistemas e Computadores, Investigação e Desenvolvimento em Lisboa e sua equipe. Concluindo as bases de dados encontradas, o ReLi (REsenha de Livros) foi criado no âmbito do projeto Anotadores Semânticos baseados em Aprendizado Ativo, do LEARN, coordenado por Ruy Milidiú (Departamento de Informática - PUC-Rio). Consiste em 1600 resenhas de livros anotadas manualmente quanto à presença de opinião sobre o livro resenhado e sua polaridade [?]. Todas as bases, já normalizadas, podem ser encontradas no apêndice A deste trabalho.

A construção da massa de treino foi feita com o intuito de melhor expressar um sentimento de uma palavra ou texto, unindo a perspectiva de diversas bases de dados, visando o melhor desempenho do algoritmo quando utilizado para classificar textos em português. Outro ponto interessante é que cada base de palavras possui um formato diferente e para incorporá-las a massa de treino do algoritmo de uma forma única foi necessário normalizá-las. Para isso, cada base foi dividida em dois arquivos: um contendo apenas termos positivos - pos.txt - e outro apenas termos negativos - neg.txt . Segue um exemplo de como uma das normalizações foi feita:

```

à-vontade,à-vontade.PoS=N;FLEX=ms;TG=HUM:N0;POL:N0=1;ANOT=MAN
abafada,abafado.PoS=Adj;FLEX=fs;TG=HUM:N0;POL:N0=-1;ANOT=JALC
abafadas,abafado.PoS=Adj;FLEX=fp;TG=HUM:N0;POL:N0=-1;ANOT=JALC
abafado,abafado.PoS=Adj;FLEX=ms;TG=HUM:N0;POL:N0=-1;ANOT=JALC
abafados,abafado.PoS=Adj;FLEX=mp;TG=HUM:N0;POL:N0=-1;ANOT=JALC
abafante,abafante.PoS=Adj;FLEX=fs|ms;TG=HUM:N0;POL:N0=-1;ANOT=MAN
abafantes,abafante.PoS=Adj;FLEX=fp|mp;TG=HUM:N0;POL:N0=-1;ANOT=MAN
abaixada,abaixado.PoS=Adj;FLEX=fs;TG=HUM:N0;POL:N0=-1;ANOT=JALC
abaixadas,abaixado.PoS=Adj;FLEX=fp;TG=HUM:N0;POL:N0=-1;ANOT=JALC
abaixado,abaixado.PoS=Adj;FLEX=ms;TG=HUM:N0;POL:N0=-1;ANOT=JALC
abaixados,abaixado.PoS=Adj;FLEX=mp;TG=HUM:N0;POL:N0=-1;ANOT=JALC
abalada,abalado.PoS=Adj;FLEX=fs;TG=HUM:N0;POL:N0=-1;ANOT=JALC
abaladas,abalado.PoS=Adj;FLEX=fp;TG=HUM:N0;POL:N0=-1;ANOT=JALC
abalado,abalado.PoS=Adj;FLEX=ms;TG=HUM:N0;POL:N0=-1;ANOT=JALC
abalados,abalado.PoS=Adj;FLEX=mp;TG=HUM:N0;POL:N0=-1;ANOT=JALC
abalizada,abalizado.PoS=Adj;FLEX=fs;TG=HUM:N0;POL:N0=1;ANOT=JALC
abalizadas,abalizado.PoS=Adj;FLEX=fp;TG=HUM:N0;POL:N0=1;ANOT=JALC
abalizado,abalizado.PoS=Adj;FLEX=ms;TG=HUM:N0;POL:N0=1;ANOT=JALC

```

Figura 3.2: Base SentiLex-PT antes da normalização

E após a divisão dos termos positivos e negativos em arquivos distintos, o resultado segue abaixo:

à-vontade	abafado
abalizado	abafante
abalroar	abaixado
abanar	abalado
abastado	abalroado
abater	abandalhado
abençoado	abandalhamento
aberto	abandonado
abnegado	abandonar
abonado	abarrotado
abonatório	abatido
aborrecer	abelhudo
abraçado	aberração
abrilhantado	aberrante
abrir a alma	aberrativo
abrir o coração	abespinhado
abrir os olhos	abestalhado
absorvente	abilolado
abstémico	abjeção
abstémio	abjecção
acatado	abjecto
aceitável	abjeto
acertado	abobado
acertar na mosca	abobalhado
	abolido
	abominador
	abominando
	abominável
	aborrecer-se
	aborrecido

(a) Lista de termos positivas

(b) Lista de termos negativas

Figura 3.3: Base SentiLex-PT normalizada, dividindo em dois arquivos distintos os termos positivos e negativos

Por último, durante os testes com a classificação surgiu a ideia de produzir uma base de dados customizada para o evento Oscar 2016. A base de dados foi elaborada seguindo o padrão estabelecido anteriormente e preenchida com palavras relevantes ao contexto do evento e classificadas manualmente pelo idealizadores deste projeto.

Leonardo DiCaprio	
Vencedor	
Primeiro oscar	
ecologia	
spotlight	
mad max	desigualdade
o regresso	estupro
tapete vermelho	racismo
favoritos	boicote
gloria pires	totalmente branca

(a) Lista de termos positivas (b) Lista de termos negativas

Figura 3.4: Amostragem da base de termos *Oscar 2016* normalizada, dividindo em dois arquivos distintos os termos positivos e negativos

3.3.5 Classificação

O primeiro passo para a classificação de um *tweet* é calibrar o algoritmo escolhido para efetuar a classificação - lembrando que para este trabalho foi utilizado o algoritmo *Naive Bayes*. Com posse das bases de palavras e termos já classificados anteriormente foi necessário treinar o algoritmo para que ele conseguisse se calibrar através das palavras e termos classificados contidos nas bases. Como já foi abordado anteriormente, para este projeto foi utilizada a biblioteca NLTK, construída como suíte de soluções para trabalhar com linguagem natural em Python. O NLTK possui diversos algoritmos de classificação disponíveis para uso, sendo o *Naive Bayes* um deles.

Por padrão, para treinar o algoritmo, o NLTK utiliza como base de dados alguns bancos de palavras genéricas disponíveis dentro da própria biblioteca. Para utilizar as bases de dados mencionadas acima como massa de treino para o algoritmo foi necessário sobrescrever os métodos de treino do classificador contidos dentro do NLTK. Após esta customização, o algoritmo passou a classificar textos utilizando as bases descritas acima como massa de treino.

Como resultado da classificação, o *Naive Bayes* retorna dois valores como resultado: o primeiro sendo a probabilidade entre zero e um do texto classificado ser positivo e outro sendo a probabilidade do mesmo ser negativo. A soma entre as duas probabilidades sempre resultam em um. Exemplo:

```
1 In[1]: from classifier import SentimentClassifier
2 In[2]: sentiment = SentimentClassifier.classify("Spotlight foi sensacional! #
        oscars2016")
3 In[3]: print sentiment
4
5 Out[1]: Sentiment({p_pos: 0.91, p_neg: 0.09, classification: "pos"})
```

Durante testes com o algoritmo de classificação *Naive Bayes* foi observado que quando o mesmo não conseguia extrair nenhuma tendência positiva ou negativa do texto a ser classificado o resultado determinava uma probabilidade igual do texto ser positivo ou negativo. Por este motivo, para facilitar as consultas e análises posteriores foi feita uma ligeira modificação para salvar o sentimento nesses casos como *"neu"* - uma alusão a palavra *"neutro"*.

Como foi dito anteriormente em na seção 3.2 (Armazenamento) antes de armazenar o documento extraído da API do Twitter são adicionadas algumas informações pertinentes ao processo de classificação. Entre elas podemos citar:

- ***normalized_text***: texto normalizado sem *stopwords* e apenas contendo o radical de cada palavra do *tweet*;
- ***classification***: classificação semântica que pode variar em positiva, neutra ou negativa;
- ***p_pos***: probabilidade do texto classificado ser positivo segundo o algoritmo;
- ***p_neg***: probabilidade do texto classificado ser negativo segundo o algoritmo;
- ***databases***: bancos de dados de palavras utilizados como massa de treino para a classificação;
- ***classified_at***: data e hora da classificação.

Estas informações foram incluídas por serem consideradas úteis durante o processo de análise dos resultados - tornando a consulta na base de dados mais simples - e durante o processo de análise do desempenho do algoritmo escolhido - mostrando por exemplo quais bases de palavras se mostraram mais eficientes para o estudo de caso conduzido durante o evento de entrega do Oscar 2016, entre outras métricas e análises que serão abordadas no Capítulo 4.

Por se tratar de algoritmo baixa complexidade matemática, o processo de classificação se mostrou surpreendentemente rápido e não se mostrou um gargalo para o processo de coleta de dados no Twitter, que apenas precisa buscar os últimos *tweets* na API do Twitter. Por conta disso, foi possível extrair os *tweets* e classificá-los imediatamente sem atrasar ou prejudicar a etapa de coleta.

Capítulo 4

Resultados e análises

Neste capítulo serão apresentados os processos e os resultados obtidos nesse trabalho. Como visto anteriormente no Capítulo 3, uma das etapas necessárias para a Análise de Sentimento é a classificação de polaridade dos *tweets*. Durante a classificação, resultado da execução do Naive Bayes, foram utilizados diferentes parâmetros que serão apresentados e discutidos durante este capítulo.

4.1 Cenários e parâmetros de teste

Durante a execução dos testes para a análise de resultados o ambiente utilizado foi:

- Sistema operacional: Linux Ubuntu 15.04
- Processador: Core i7
- Memória: 8GB
- Quantidade de *tweets*: 141.798

4.2 Técnicas

Durante a realização do trabalho foram utilizadas duas técnicas para maximizar a performance do modelo. As técnicas foram aplicadas nas bases utilizadas para a classificação e nos dados coletados de acordo com os testes discutidos ao longo do capítulo

Palavra	Stemização
boate	boat
boates	boat
boca	boc
bocados	boc

Tabela 4.1: Exemplo de stemização

4.2.1 Stemming

A técnica de *stemming*, conhecido em português como stemização, consiste na redução de um termo ao seu radical, removendo as desinências, afixos, e vogais temáticas. Com sua utilização, os termos derivados de um mesmo radical serão contabilizados como um único termo [?].

4.2.2 Stopwords

A técnica de *stopwords*, como citado anteriormente, consiste na remoção de palavras "vazias", como artigos, preposições e interjeições, que não agregam valor a análise realizada. Além disso, essas palavras podem ser configuráveis dependendo do domínio do seu estudo [?].

4.3 Desenvolvimento do modelo de análise

O primeiro teste realizado para a classificação da base obteve o seguinte resultado:

1º Teste	
Bases usadas	Técnicas usadas
Sentilex	Stopwords
PUC	Stemming
ReLi	
Resultado	
Positivo	17.350
Negativo	15.517
Neutro	108.931

Tabela 4.2: 1º teste

Como mostra a tabela 4.2 é visto quais as bases utilizadas, nesse caso, ReLi , PUC e

Sentilex, as técnicas utilizadas nesse teste, *Stopwords* e *Stemming*, e o resultado que de 141.798 *tweets*, 17.350 foram positivos, 15.517 negativos e 108.931 neutros.

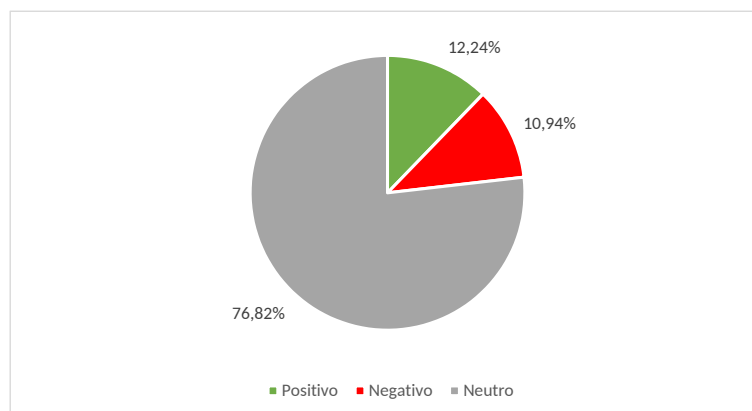


Figura 4.1: Quantidade de tweets separados por polaridade do teste 1. Fonte: Própria

Nota-se que a quantidade de *tweets* neutros é elevada, evidenciando que o modelo ainda tem dificuldade de definir a polaridade do texto. Com base nos resultados apresentados foram realizadas as seguintes mudanças visando diminuir a ocorrência de dados neutros.

2º Teste	
Bases usadas	Técnicas usadas
Sentilex-Stem	Stopwords
PUC-Stem	Stemming
ReLi-Stem	
Resultado	
Positivo	49.263
Negativo	35.079
Neutro	57.456

Tabela 4.3: 2º teste

Como mostra a tabela 4.3 as mesmas bases foram utilizadas, porém com a aplicação da técnica de *stemming*, como PUC-Stem e Sentilex-Stem. Além de utilizar a mesma técnica nas *stopwprds*. O resultado de 141.798 *tweets*, 49.263 foram positivos, 35.079 negativos e 57.456 neutros.

Analisando o 2º teste é visto que a quantidade de neutro diminuiu consideravelmente, apenas aplicando a técnica de *stemming* nas bases de palavras.

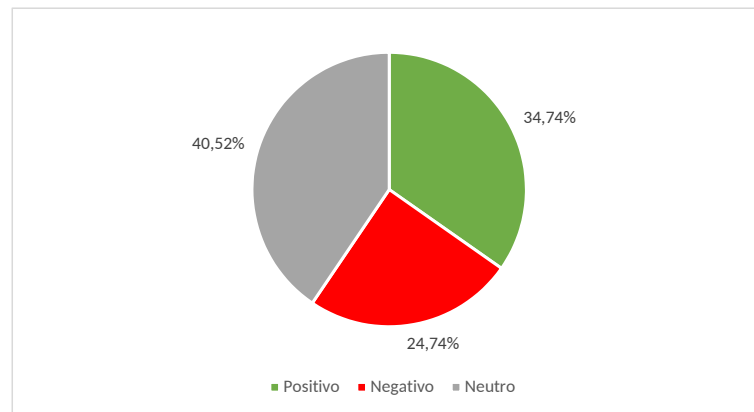


Figura 4.2: Quantidade de tweets separados por polaridade do teste 2. Fonte: Própria

Ainda buscando a diminuição de dados neutros foi criada uma base de palavras mais próxima do domínio que esse trabalho propõe com a base chamada Oscar2016, essa base contém palavras relevantes ao evento, gerando o seguinte resultado.

3º Teste	
Bases usadas	Técnicas usadas
Oscar2016	<i>Stopwords</i>
Resultado	
Positivo	47.450
Negativo	7.210
Neutro	87.138

Tabela 4.4: 3º teste

Analisando a tabela 4.4 é visto que apenas uma base mais especializada no domínio não consegue diminuir a quantidade de neutros. O resultado de 141.798 *tweets*, 47.450 foram positivos, 7.210 negativos e 87.138 neutros.

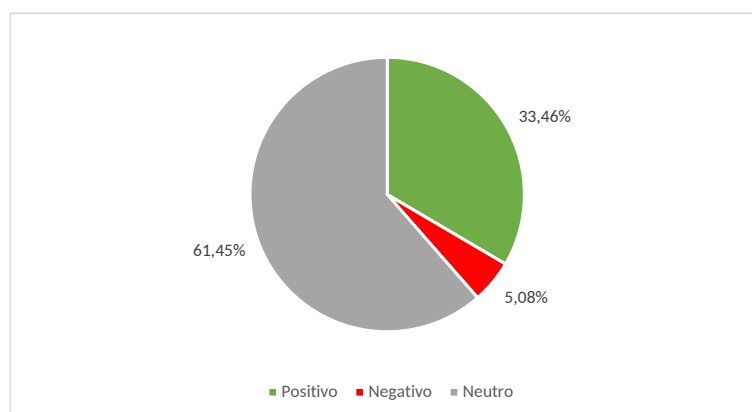


Figura 4.3: Quantidade de tweets separados por polaridade do teste 3. Fonte: Própria

No 4º teste foi adicionada a base criada, Oscar2016 com as bases genéricas, Sentilex, PUC e Reli gerando o seguinte resultado:

4º Teste	
Bases usadas	Técnicas usadas
Oscar2016	Stopwords
SentiLex	Stemming
PUC	
ReLi	
Resultado	
Positivo	69.070
Negativo	33.461
Neutro	39.267

Tabela 4.5: 4º teste

Analisando a tabela 4.5 é visto que nesse teste foi obtido a maior taxa de diminuição de dados neutros. O resultado de 141.798 *tweets*, 69.070 foram positivos, 33.461 negativos e 39.267 neutros.

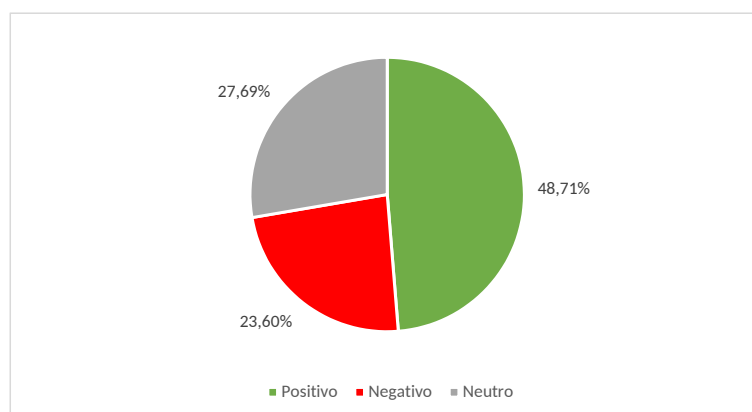


Figura 4.4: Quantidade de tweets separados por polaridade do teste 4. Fonte: Própria

Segue abaixo um comparativo dos testes.

	Teste 1	Teste 2	Teste 3	Teste 4
Positivo	15.517	49.263	47.450	69.070
Negativo	17.350	35.079	7210	33.461
Neutro	108.931	57.456	87.138	39.267

Tabela 4.6: Comparando testes

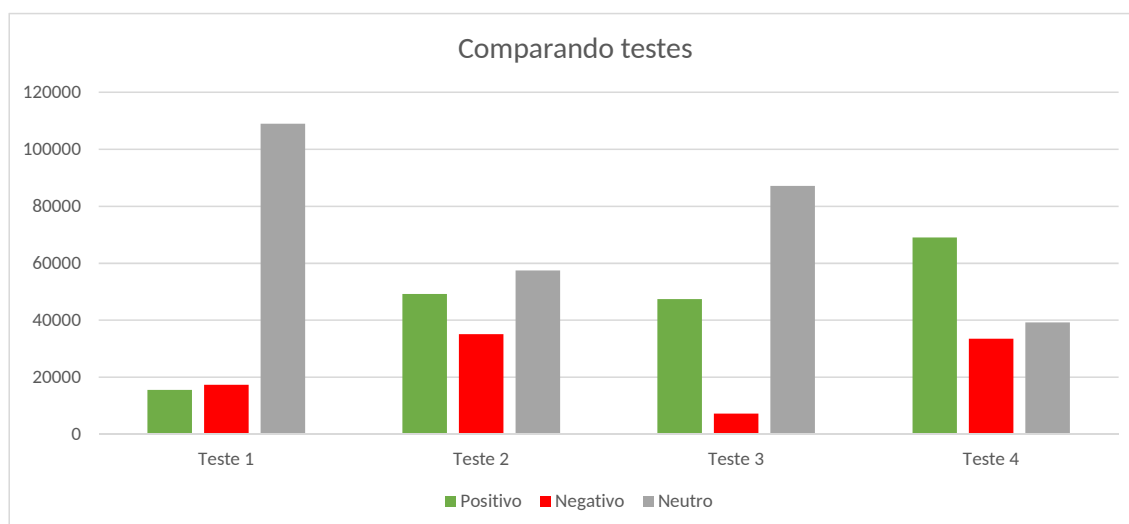


Figura 4.5: Gráfico de comparação dos testes

Com o comparativo analisado na tabela 4.6 foram escolhidas as configurações utilizadas no teste 4 para a realização da análise na base capturada, devido a menor presença de dados neutros classificados na base comprovando que as técnicas de *stemming* e *stopwords* foram válidas e relevantes para a sustentação modelo.

4.4 Resultados

Para a análise dos resultados é necessário estabelecer premissas. O evento Oscar 2016 foi a 88.^a cerimônia de entrega dos *Academy Awards* em *Los Angeles*, Estados Unidos. O evento teve duração de aproximadamente 4 horas com seu início na noite do dia 28 de fevereiro, às 20:00 horário de Brasília no tapete vermelho com a entrada dos artistas e convidados. Seu encerramento às 2:00 do horário de Brasília do dia 29 de Fevereiro com a premiação do melhor filme. Os marcos do evento relevantes para o estudo:

- Apresentação do David Grohe em memória póstuma
- Apresentação da Lady Gaga
- Discurso do Vice Presidente Americano
- Melhor atriz
- Melhor ator
- Melhor filme

Todos os eventos ocorridos no Oscar 2016 são listados em ordem cronológica no apêndice B

O gráfico 4.6 mostra a curva da quantidade de *tweets* em relação ao tempo dividido pela polaridade. Nele é visto um pico no começo do evento, durante o tapete vermelho, nesse período foi visto as maiores ocorrências de especulações sobre os ganhadores e indicados além das reações dos internautas com as entradas de seus artistas favoritos.

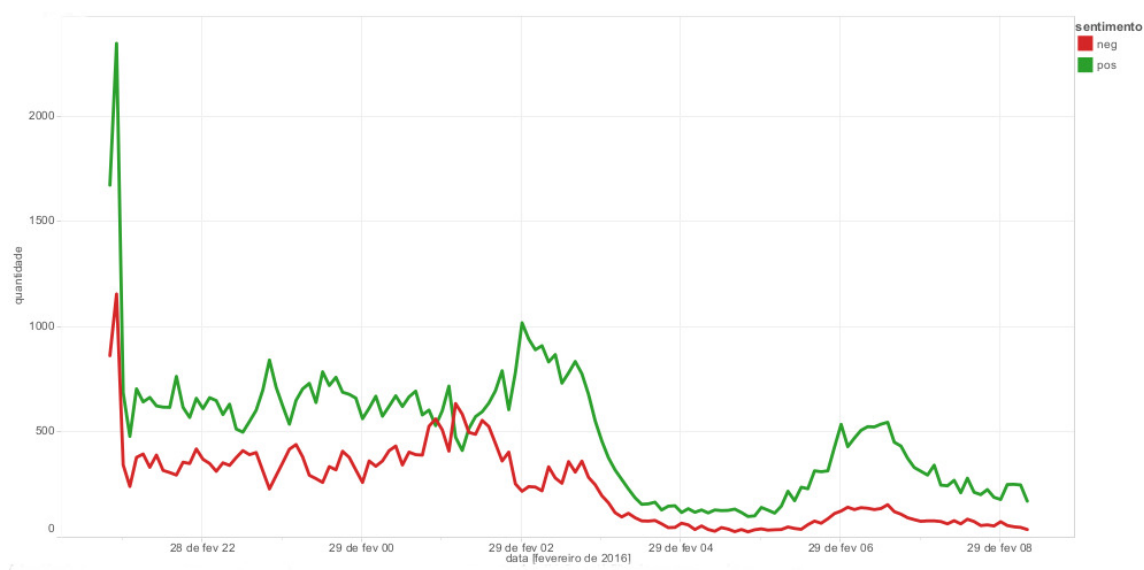


Figura 4.6: Quantidade de tweets por tempo e polaridade. Fonte:Própria

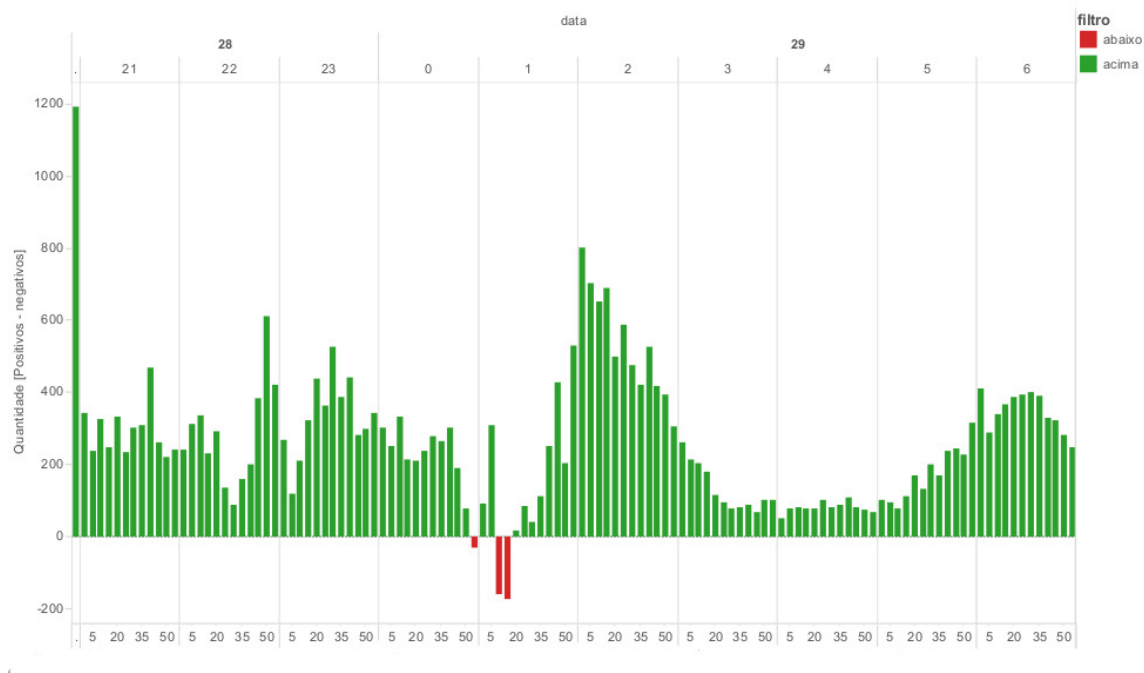


Figura 4.7: Quantidade de tweets positivos diminuído pelos negativos pelo tempo. Fonte:Própria

O gráfico 4.7 foi gerado a partir da diminuição dos *tweets* negativos pelos *tweets* positivos vistos no gráfico 4.6 produzindo uma visão micro do dados coletados do evento. Analisando o gráfico 4.6, que possui uma visão macro, junto com o gráfico 4.7, com a visão micro, é notado que em dois momentos a quantidade de *tweets* negativos sobrepõem a quantidade de positivos. No primeiro momento, é durante o *show* do Dave Grohl onde

é feita uma homenagem aos falecidos no mundo do cinema em 2015 onde a sensação de luto ficou evidente durante a apresentação, e de acordo com [?], é a reação à perda de objeto ou pessoa. Como afeto, o luto aproxima-se do humor depressivo. Com isso foi corroborado o aumento do sentimento negativo nessa parte do evento. Outro momento onde ocorre o mesmo efeito é durante a apresentação da Lady Gaga onde o tema do filme que a música faz parte da trilha é sobre os estupros em faculdades americanas, e mais uma vez esse aumento negativo faz entender o tema pesado e sensível para as pessoas que sofreram e sofrem disso. Excluindo o pico no começo dos gráficos os maiores picos de positivos são presenciados as 2:00, vindo de uma crescente no final das 1:00 da manhã do dia 29 isso se deve as categorias mais aguardadas do evento que são:

- Melhor atriz
- Melhor ator
- Melhor filme

O maior destaque foi a premiação de melhor ator e a vitória de Leonardo DiCaprio, que com seus 41 anos de idade e 26 de carreira, nunca havia levado um prêmio de melhor ator da Academia. Essa vitória reflete nos gráficos de picos de positivos as 2:00 da manhã onde a repercussão da sua vitória é espelhada pelo twitter chegando aos *trending topics* mundias na rede social.

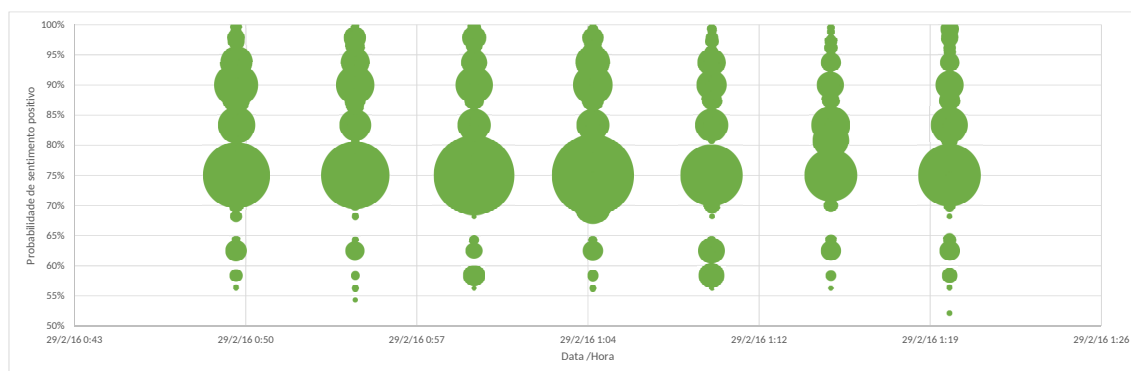


Figura 4.8: Gráfico de bolhas referente aos eventos ocorridos entre 00:50 e 01:20 de probabilidade positiva

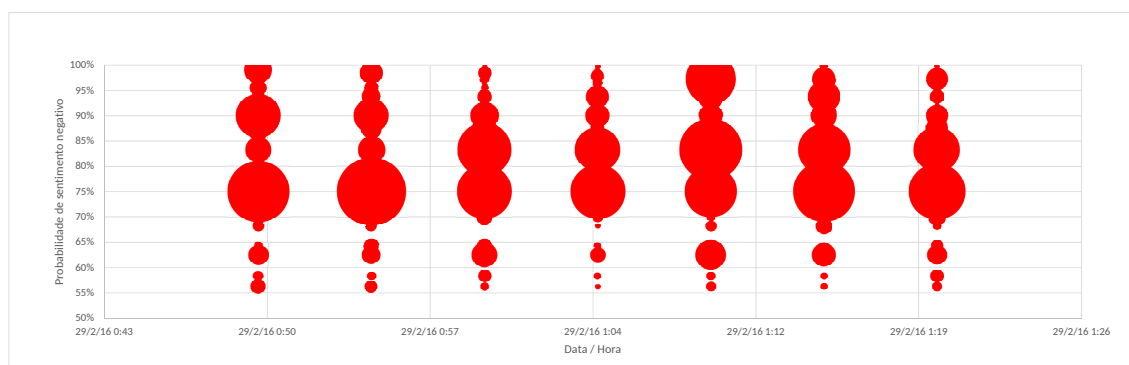


Figura 4.9: Gráfico de bolhas referente aos eventos ocorridos entre 00:50 e 01:20 de probabilidade negativa

Mediante os resultados encontrados nos gráficos 4.8 e 4.9, o algoritmo utilizado mostra que na faixa de 65% a 90% , foi a região onde sua classificação obteve a maior quantidade de resultados tanto para *tweets* positivos quanto para negativos. É visto que próximo ao marco de 1:12 a quantidade de *tweets* negativos é maior , confirmando a análise feita na figura 4.7

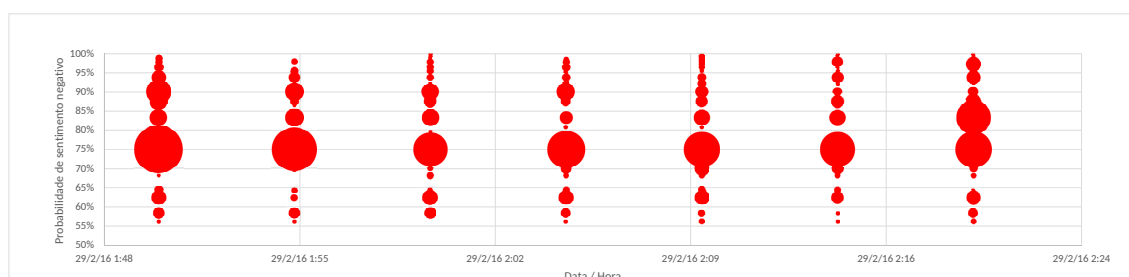


Figura 4.10: Gráfico de bolhas referente aos eventos ocorridos entre 01:50 e 02:20 de probabilidade positiva

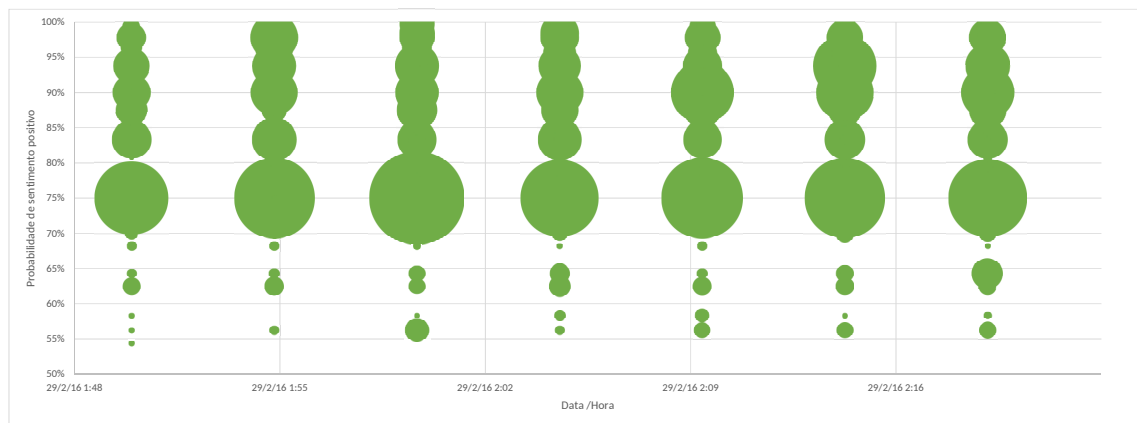


Figura 4.11: Gráfico de bolhas referente aos eventos ocorridos entre 01:50 e 02:20 de probabilidade negativa

Já analisando os gráficos 4.10 e 4.11, percebe-se que houve mudanças na faixa de probabilidade. Uma delas foi nos *tweets* classificados como negativos onde a faixa ficou entre 70% e 80% e a quantidade, o tamanho das bolhas, consideravelmente menor. Já a outra mudança ocorre no marco de 2:09 na probabilidade de positivos, onde a faixa ultrapassa os 90% corroborando a análise feita na figura 4.7 onde é visto os picos de *tweets* positivos.

Capítulo 5

Conclusão

Este trabalho identificou e abordou a importância que a mineração de opinião possui nos dias atuais, além das oportunidades que este novo campo de estudo abre com a sua descoberta: tanto na área acadêmica quanto para exploração comercial e estratégica e assim como propôs, desenvolveu uma solução capaz de extrair mensagens escritas em língua portuguesa de uma *hashtag* do Twitter e classificá-las em positivas, neutras ou negativas, tornando possível fazer análises de mineração de opinião que são capazes de interpretar como um universo de usuários se sente em relação a temática estudada. Sem dúvida, o grande desafio foi mostrar que é viável aplicar uma solução de mineração de opinião em língua portuguesa. Com base no que foi apresentado, a proposta cumpriu os seguintes objetivos levantados:

- Utilização do Twitter como plataforma de dados;
- Construção um processo escalável capaz de extrair mensagens de uma *hashtag* do Twitter;
- Utilização bases de palavras classificadas disponíveis na literatura sobre o tema;
- Aplicação de técnicas de normalização de texto visando um maior desempenho ao classificar;
- Implementação do algoritmo de *Naive Bayes* para classificar as mensagens em positivas, neutras ou negativas;
- Conduzir a classificação em mensagens que foram escritas em língua portuguesa;
- Aplicação do conhecimento adquirido durante este trabalho em um estudo de caso utilizando a cerimônia do Oscars 2016;

- Demonstração dos cenários de teste e seus impactos no resultado final;
- Análise dos resultados obtidos através de gráficos que permitem interpretar os sentimentos dos usuários antes, durante e depois do evento, sob diversas perspectivas.

A análise dos resultados mostrou que por existir uma ansiedade muito grande vinculada aos instantes que antecedem o evento existe uma grande concentração de sentimentos positivos atrelados a este momento. Outro fato que ficou evidente é que os momentos mais populares do evento, como o anúncio das categorias mais aguardadas como: melhor ator, melhor atriz e melhor filme são as mais esperadas e com a confirmação da vitória dos favoritos, houve um pico de mensagens positivas por volta do horário do anúncio dos vencedores destas categorias. Outros dois momentos chamaram atenção pela carga de sentimentos negativos: o show de David Grohl em homenagem aos falecidos no mundo do cinema em 2015 onde a sensação de luto ficou evidente e a apresentação de Lady Gaga, onde o tema do filme que a música faz parte da trilha sonora aborda estupros em faculdades americanas, tema que desperta bastante tristeza nas pessoas que sofreram este tipo de trauma ou que empatizam de alguma forma com as vítimas deste crime.

5.1 Trabalhos Futuros

Como visto neste trabalho, o algoritmo utilizado para gerar os resultados das classificações afim de realizar a análise de sentimento foi o *Naive Bayes*. Para trabalhos futuros pode-se utilizar outros algoritmos como: Árvore de Decisão, Regressão Logística, *Maximum Entropy Model*. Podendo ser usados somente o próprio algoritmo ou realizando um trabalho comparativo entre eles, descobrindo qual algoritmo gera melhores resultados.

Outro método que pode ser utilizado é a aplicação de *n-grams*, para averiguar o melhor desempenho do modelo. Com isso poderia-se realizar testes de acurácia, que não foi o enfoque deste trabalho. Pode-se também utilizar outras fontes de dados, não só o Twitter, como: Facebook, fóruns, comentários de sites ,entre outros. Mais um fator relevante são as bases, que nesse trabalho foram utilizadas três bases genéricas e uma pequena base personalizada, então como trabalho futuro propõem-se utilizar apenas bases personalizadas e referentes ao domínio.

Durante as pesquisas para esse trabalho foi encontrado um *toolkit* para a técnica de *word embedding*, conhecido como *word2vec*, criado em 2013 por uma equipe do Google. Esse tipo de técnica se baseia em redes neurais, que difere do algoritmo probabilístico

utilizado nesse trabalho. Sendo essa uma outra proposta para trabalhos futuros.

Entre os testes realizados no capítulo 4 foi presenciado um fator instigante. A duração dos testes, que podem ser encontrados no apêndice B.1, foi notado um crescimento anômalo, entretanto não foi discutido a causa de tal crescimento, então é proposto um estudo onde se discute tal anomalia.

Por fim, pretende-se fazer o máximo para desenvolver o processamento de linguagem natural para a língua portuguesa. Onde técnicas e métodos serão comparadas visando o melhor desempenho com uma análise mais profunda e a automação da inteligência artificial para o aproveitamento desse campo para o ser humano em suas tomadas de decisão.

Referências

- [1] G. Miner, *Practical text mining and statistical analysis for non-structured text data applications*. Academic Press, 2012.
- [2] H. M. M. Lastres, S. Albagli, and C. A. K. Passos, *Informação e globalização na era do conhecimento*. Campus Rio de Janeiro, 1999.
- [3] H. J. C. Gomes, “Text mining: análise de sentimentos na classificação de notícias,” Ph.D. dissertation, 2013.
- [4] A. Pak and P. Paroubek, “Twitter as a corpus for sentiment analysis and opinion mining,” in *LREc*, vol. 10, 2010, pp. 1320–1326.
- [5] D. Santos, “O projecto processamento computacional do português: Balanço e perspectivas,” *quot; In Maria das Graças Volpe Nunes (ed) V Encontro para o processamento computacional da língua portuguesa escrita e falada (PROPOR 2000)(Atibaia SP 19-22 de Novembro de 2000) São Paulo: ICMC/USP, 2000.*
- [6] P. L. Tortella and J. M. A. Coello, “Análise de sentimentos em mídias sociais.”
- [7] G. A. Rodrigues Barbosa, I. S. Silva, M. Zaki, W. Meira Jr, R. O. Prates, and A. Veloso, “Characterizing the effectiveness of twitter hashtags to detect and track online population sentiment,” in *CHI’12 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2012, pp. 2621–2626.
- [8] T. Franca and J. Oliveira, “Análise de sentimento de tweets relacionados aos protestos que ocorreram no brasil entre junho e agosto de 2013,” in *III Brazilian Workshop on Social Network Analysis and Mining (BraSNAM)*, 2014.
- [9] J. A. CARVALHO FILHO, “Mineração de textos: Análise de sentimento utilizando tweets referentes à copa do mundo 2014,” 2014.
- [10] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, “From data mining to knowledge discovery in databases,” *AI magazine*, vol. 17, no. 3, p. 37, 1996.
- [11] K. R. Scherer and M. R. Zentner, “Emotional effects of music: Production rules,” *Music and emotion: Theory and research*, pp. 361–392, 2001.
- [12] F. L. d. Santos, “Mineração de opinião em textos opinativos utilizando algoritmos de classificação,” 2014.
- [13] C. A. S. R. et al., “Mineração de opinião / análise de sentimento.” [Online]. Available: <http://www.inf.ufsc.br/~alvares/INE5644/MineracaoOpinioao.pdf>

- [14] G. Villela and P. A. Mendes, “Finanças comportamentais: O impacto da razão e da emoção no processo decisório em investimentos no mercado financeiro brasileiro,” *Revista de Administração da FATEA*, vol. 6, no. 6, pp. 81–92, 2013.
- [15] [Online]. Available: "<http://www.whatdoestheinternetthink.net>"
- [16] (2016, Maio) Twitter company. [Online]. Available: <https://about.twitter.com/company>
- [17] (2016, Maio) Twitter company. [Online]. Available: "<https://dev.twitter.com/overview/api/counting-characters>"
- [18] (2016, Abril) Dmr stats. [Online]. Available: <http://expandedramblings.com/index.php/march-2013-by-the-numbers-a-few-amazing-twitter-stats/>
- [19] M. Waite, *Paperback Oxford English dictionary*. Oxford University Press, 2012.
- [20] P. N. Howard, A. Duffy, D. Freelon, M. M. Hussain, W. Mari, and M. Mazaid, “Opening closed regimes: what was the role of social media during the arab spring?” *Available at SSRN 2595096*, 2011.
- [21] S. Wasserman and J. Galaskiewicz, *Advances in social network analysis: Research in the social and behavioral sciences*. Sage Publications, 1994, vol. 171.
- [22] (2016, Maio) Twitter company. [Online]. Available: "<https://dev.twitter.com/overview/documentation>"
- [23] R. Fielding, “Architectural styles and the design of network-based software architectures.” [Online]. Available: http://www.ics.uci.edu/~fielding/pubs/dissertation/rest_arch_style.htm
- [24] S. Weber, *The success of open source*. Cambridge Univ Press, 2004, vol. 368.
- [25] M. A. Covington, *Natural language processing for Prolog programmers*. Prentice Hall Englewood Cliffs (NJ), 1994.
- [26] A. M. Turing, “Computing machinery and intelligence,” *Mind*, vol. 59, no. 236, pp. 433–460, 1950.
- [27] J. D. Rennie, L. Shih, J. Teevan, D. R. Karger *et al.*, “Tackling the poor assumptions of naive bayes text classifiers,” in *ICML*, vol. 3. Washington DC), 2003, pp. 616–623.
- [28] I. Androutsopoulos, J. Koutsias, K. V. Chandrinou, G. Paliouras, and C. D. Spyropoulos, “An evaluation of naive bayesian anti-spam filtering,” *arXiv preprint cs/0006013*, 2000.
- [29] G. Lucca, I. A. Pereira, A. Prisco, and E. N. Borges, “Uma implementação do algoritmo naïve bayes para classificação de texto,” 2013.
- [30] (2016, Maio) Twitter company. [Online]. Available: <https://apps.twitter.com/>
- [31] [Online]. Available: "<http://oauth.net/>"

- [32] (2016, Maio) Twitter company. [Online]. Available: "<https://dev.twitter.com/rest/public/rate-limiting>"
- [33] K. Jeffay, "The real-time producer/consumer paradigm: A paradigm for the construction of efficient, predictable real-time systems," in *Proceedings of the 1993 ACM/SIGAPP symposium on Applied computing: states of the art and practice*. ACM, 1993, pp. 796–804.
- [34] (2016, Maio) Twitter company. [Online]. Available: "<https://dev.twitter.com/rest/public/search>"
- [35] (2016, Maio) Twitter company. [Online]. Available: "<https://dev.twitter.com/rest/public/timelines>"
- [36] N. T. Bhuvan and M. S. Elayidom, "A technical insight on the new generation databases: Nosql," *International Journal of Computer Applications*, vol. 121, no. 7, 2015.
- [37] "voldemort project. [Online]. Available: "<http://www.project-voldemort.com/voldemort/>"
- [38] "mongodb. [Online]. Available: "<https://www.mongodb.com/>"
- [39] "tokyo cabinet: a modern implementation of dbm. [Online]. Available: "<http://fallabs.com/tokyocabinet/>"
- [40] "couchdb". [Online]. Available: "<https://couchdb.apache.org/>"
- [41] F. Chang, J. Dean, S. Ghemawat, W. C. Hsieh, D. A. Wallach, M. Burrows, T. Chandra, A. Fikes, and R. E. Gruber, "Bigtable: A distributed storage system for structured data," *ACM Transactions on Computer Systems (TOCS)*, vol. 26, no. 2, p. 4, 2008.
- [42] G. DeCandia, D. Hastorun, M. Jampani, G. Kakulapati, A. Lakshman, A. Pilchin, S. Sivasubramanian, P. Vosshall, and W. Vogels, "Dynamo: amazon's highly available key-value store," *ACM SIGOPS Operating Systems Review*, vol. 41, no. 6, pp. 205–220, 2007.
- [43] "natural language toolkit 3.0 documentation". [Online]. Available: "<http://www.nltk.org/>"
- [44] "examples for portuguese processing". [Online]. Available: "http://www.nltk.org/howto/portuguese_en.html"
- [45] "linguateca". [Online]. Available: "<http://www.linguateca.pt/>"
- [46] "visual interactive syntax learning". [Online]. Available: "<http://visl.sdu.dk/>"
- [47] C. Freitas, "Sobre a construção de um léxico da afetividade para o processamento computacional do português," *Revista Brasileira de Linguística Aplicada*, vol. 13, no. 4, pp. 1013–1059, 2013.

- [48] P. C. "Mário J. Silva and L. Sarmiento", "lecture notes in computer science", in *"Building a Sentiment Lexicon for Social Judgement Mining"*, "International Conference on Computational Processing of the Portuguese Language (PROPOR)". "Springer", "2012", pp. "218–228".
- [49] R. Milidião. "reli (resenha de livros)". [Online]. Available: "<http://www.linguateca.pt/Repositorio/ReLi/>"
- [50] V. M. Orenço and C. Huyck, "A stemming algorithm for the portuguese language," in *String Processing and Information Retrieval, 2001. SPIRE 2001. Proceedings. Eighth International Symposium on*, Nov 2001, pp. 186–193.
- [51] A. Rajaraman and J. D. Ullman, *Mining of Massive Datasets*. Cambridge University Press, Oct 2011. [Online]. Available: <https://www.cambridge.org/core/books/mining-of-massive-datasets/A06D57FC616AE3FD10007D89E73F8B92>
- [52] S. Freud, "Conferências introdutórias sobre psicanálise," in *Conferência XXIII*, 1908.

APÊNDICE A - Bases de texto utilizadas

A.1 *PUC*

A.1.1 Termos negativos

anticlímax besteira bobagem boçal bosta chatice clichê decepção defeito demorar desastre desconforto desesperança desgraça desprazer droga engano erro exagero exaustão excesso falhar falho falta furo imaturidade incômodo inexperiente lixar meleca melodrama melodramático meloso mundo nojo passividade pena pieguice porca porcaria problema sacrifício sentimentalismo superficialidade vergonha abandonar aborrecer arrepender assustar aterrorizar atormentar atrapalhar cansar chocar complicar decepcionar deprimir desanimar desgastar desistir desmerecer destorcer detestar dificultar distorcer empacar enfraquecer enganar engolir errar estragar estressar exagerar faltar frustrar incomodar irritar largar limitar odiar pecar perder prolongar revoltar aborrecente água-com-açúcar anacrônico besta bestar bizarro bobo burro cansativo chato chocante chulo clichê confuso construir decepcionante defeituoso deplorável depressivo deprimente desagradável desconexo desgastante desinteressante desnecessário desprezível difícil dispensável doentio egoísta enfadonho enjoado enjoativo entediante esdrúxulo estereotipado estranho falsar falso fraco frio frustrante fútil graça horrível horrorosa idiota idiotice imaturo impaciente impressionar incompreensível inconsistente ingênuo injustificável insuportável interminável inútil irritante lamentável maçante machista mal manjado massa mau mediano médio meloso mero modinha morno negativo normal obsessivo oco opressivo patético pesaroso péssimo piegas piorar plana pobre pomposo preguiçoso previsível puritano rasa raso razoável reacionário reclamo repa repetitivo repulsivo revoltante ridículo ruim sem simplista sofrível superficial surreal tédio tedioso tenda tongo tosco triste vazio violento volúvel estático idêntico lento linear

A.1.2 Termos positivos

alegria astúcia atemporalidade beleza bem-estar best-seller brilho carisma clímax companheirismo compreensão deleite densidade diferencial diversão doçura emoção empatia entusiasmo envolvente êxtase fã facilidade fascinação favorito fluência fluidez genialidade gostinho grandeza inteligência ironia magnetismo mérito obra-prima originalidade paixão perfeição prazer predileto preferido primor sensibilidade sufocamento talento taquicardia ternura top virtude adorar agradar amadurecer amar animar apaixonar apegar apreciar aprender arrebatador arrebatador atrair cativar conquistar curtir deliciar destacar-se divertir emocionar emplacar empolgar encantar enternecer entreter facilitar fascinar fluir gostar identificar inovar recomendar sensibilizar simpatizar valorizar vibrar viciar apavorar brilhar comover devorarprender rir saborear surpreender tocar admirável adorável agradável alegre alucinante apaixonante ardente arrasador arrebatador astuto atemporal aterrador atraente atual atuar autêntico avassalador bacana bárbaro batuta belíssimo belo bem bem-humorado bom bonito brilhante carinhoso carismático cativante cativar certo claro classe clássico coerente comédia cômico comovente compreensível comum consistente contagiante contente convincente corajoso criativo crivar curto decente delicado delicioso demais denso desafiador desprezível devaneio digno direto distrativo divertido divino doce duro elegante eletrizante emocionante empolgante encantador encantável engenhoso engraçado enigmático enriquecedor envolvente especial esperto espetacular espirituoso esplêndido essa estonteante excelente excepcional excitante exemplar extraordinário fabuloso face fácil fama fantástico fascinante fascinate favorito feliz fluente fluido foder fofo forte frenético fundamental genial gostoso gotoso grande grandioso gratificante grato hilário humano ideal incompar impactante impagável impar ímpar impecável imperdível importante impressionante inabalável inacreditável incrível indescritível indispensável inédito inesquecível infalível inovador inquietante inquieto inspirador instigante instigar inteligente intenso interessante íntimo intrigante irado irônico irresistível jovial legal legítimo leve liça light linda lindo lírico mágico magistral magnético magnificar magnífico majestoso maravilha maravilhoso marcante massa máximo melífluo memorável minucioso misterioso moderno notável novo obra ode original palpável passável perfeito pertinente plausível poderoso poético pontual positivo precioso predileto primo primoroso profético profundo pungente rápido raro razoável real realista realístico recente recomendar recomendável reflexivo relaxante relevante rico sagaz seco sedutor seguro sensacional sensível sensual sentimental sério sexy significativo simples sincero singelo singular sublime subversivo surpreso surreal teres tocante tranquilo translúcido último válido verdadeiro verossímil vertiginoso viciante vicioso vigoroso visceral vital voraz arrasador aterrorizante

assustador angustiante compulsivo frenético imprevisível surpreendente sufocante perturbador sombrio

A.2 *ReLi*

A.2.1 Termos negativos

água com açúcar água-com-açúcar cabeça-dura carregar nas tintas dar náuseas dar raiva dar sono deixar a desejar embrulhar o estômago esperar mais forçar a barra mala sem alça melhor esquecer não aturar não engolir não prestar não se enxergar perder o fôlego picolé de chuchu pra criancinha quebrar o encanto sem graça sem sal sem-noção ser um balde de água fria ser um saco tempo perdido perder tempo perda de tempo aborrecente anacrônico besta bizarro bobo burro cansativo chato chocante chulo clichê confuso decepçionante defeituoso deplorável depressivo deprimente desagradável desconexo desgastante desinteressante desnecessário desprezível difícil dispensável doentio egoísta enfadonho enjoado enjoativo entediante esdrúxulo estereotipado estranho falso fraco frio frustrante fútil horrível horrorosa idiota idiotice imaturo impaciente incompreensível inconsistente ingênuo injustificável insuportável interminável inútil irritante lamentável maçante machista mal manjado mau mediano médio meloso mero modinha morno negativo normal obsessivo oco opressivo pesaroso péssimo piegas piorar plana pobre pomposo preguiçoso previsível puritano raso razoável reacionário repetitivo repulsivo revoltante ridículo ruim simplista sofrível superficial surreal tédio tedioso tenda tongo tosco triste vazio violento volúvel estático idêntico lento linear anticlímax besteira bobagem boçal bosta chatice clichê decepção defeito demorar desastre desconforto desesperança desgraça desprazer droga engano erro exagero exaustão excesso falhar falho falta furo imaturidade incômodo inexperiente lixar meleca melodrama melodramático meloso mundo nojo passividade pena pieguice porca porcaria problema sacrifício sentimentalismo superficialidade vergonha abandonar aborrecer arrepender assustar aterrorizar atormentar atrapalhar cansar chocar complicar decepçionar deprimir desanimar desgastar desistir desmerecer destorcer detestar dificultar distorcer empacar enfraquecer enganar engolir errar estragar estressar exagerar faltar frustrar incomodar irritar largar limitar odiar pecar perder prolongar revoltar

A.2.2 Termos positivos

abrir a mente abrir a minha mente amor à primeira vista arrancar lágrimas arrancar risadas arrancar gargalhadas capturar a atenção chamar a atenção com chave de ouro dar

certo de corpo e alma de encher os olhos de fazer inveja de tirar o fôlego deixar de boca aberta estar louco para faltar adjetivos faltar palavras fazer pensar fora de série gostinho de quero mais grata surpresa ir a nocaute lindo de morrer morrer de rir não ter adjetivos para pra ninguém botar defeito ser show ser um show ser um achado ser um tapa na cara soco na boca de o estômago ter fôlego tudo de bom valer cada N valer a pena admirável adorável agradável alegre alucinante apaixonante ardente arrasador arrebatador astuto atemporal aterrador atraente atual atuar autêntico avassalador bacana bárbaro batuta belíssimo belo bem bem-humorado bom bonito brilhante carinhoso carismático cativante cativar certo claro classe clássico coerente comédia cômico comovente compreensível comum consistente contagiante contente convincente corajoso criativo crivar curto decente delicado delicioso demais denso desafiador despretenso devanear digno direto distrativo divertido divino doce duro elegante eletrizante emocionante empolgante encantador encantável engenhoso engraçado enigmático enriquecedor envolvente especial esperto espetacular espirituoso esplêndido essa estonteante excelente excepcional excitante exemplar extraordinário fabuloso face fácil fama fantástico fascinante fascinate favorito feliz fluente fluido foda fofo forte frenético fundamental genial gostoso gotoso grande grandioso gratificante grato hilário humano ideal incompar impactante impagável impar ímpar impecável imperdível importante impressionante inabalável inacreditável incrível indescritível indispensável inédito inesquecível infalível inovador inquietante inquieto inspirador instigante instingar inteligente intenso interessante íntimo intrigante irado irônico irresistível jovial legal legítimo leve liça light linda lindo lírico mágico magistral magnético magnificar magnífico majestoso maravilha maravilhoso marcante massa máximo melífero memorável minucioso misterioso moderno notável novo obra ode original palpável passável perfeito pertinente plausível poderoso poético pontual positivo precioso predileto primo primoroso profético profundo pungente rápido raro razoável real realista realístico recente recomendável reflexivo relaxante relevante rico sagaz seco sedutor seguro sensacional sensível sensual sentimental sério sexy significativo simples sincero singelo singular sublime subversivo surpreso surreal teres tocante tranquilo translúcido último válido verdadeiro verossímil vertiginoso viciante vicioso vigoroso visceral vital voraz arrasador aterrorizante assustador angustiante compulsivo frenético imprevisível; surpreendente sufocante perturbador sombrio alegria astúcia atemporalidade beleza bem-estar best-seller brilho carisma clímax companheirismo compreensão deleite densidade diferencial diversão doçura emoção empatia entusiasmo envolvente êxtase fã facilidade fascinação favorito fluência fluidez genialidade gostinho grandeza inteligência ironia magnetismo mérito obra-prima originalidade paixão perfeição prazer predileto preferido primor sensibilidade sufocamento talento

taquicardia ternura top virtude adorar agradar amadurecer amar animar apaixonar apegar apreciar aprender arrebatado arrebatado atrair cativar conquistar curtir deliciar destacar-se divertir emocionar emplacar empolgar encantar enternecer entreter facilitar fascinar fluir gostar identificar impressionar inovar recomendar sensibilizar simpatizar valorizar vibrar viciar apavorar brilhar comover devorar prender rir saborear surpreender tocar

A.3 *SentiLex*

A.3.1 Termos negativos

abafado abafante abaixado abalado abalroado abandalhado abandalhamento abandonado abandonar abarrotado abatido abelhudo aberração aberrante aberrativo abespinhado abestalhado abilolado abjeção abjecção abjecto abjeto abobado abobalhado abolido abominador abominando abominável aborrecer-se aborrecido abortado abrupto abrutalhado absentista abstracto abstraído abstrato abstrusidade abstruso absurdo abulia abúlico abusado abusador abusar abusivo acabado acabadote acabrunhado acaciano açambarcador acanhado acanhamento acanhar-se acéfalo acerado acerbidade acerbo acérrimo achacado achatado achincalhar acidentado aciganado acintoso acirrado acobardado acobardar-se acobertado acocado açoitado acomodadiço acomodar-se acordar com o rabo para o ar acordar com os pés de fora acorrentado acorrentar-se acossado açougueiro acre acriançado acrítico aculturado acusado acutilante adiposo admoestador adoentado adormecido aduldor adulterador adulterino adultério adúltero adverso aéreo afanado afasia afásico afectado afeminado afetado afiado aflição afligido afligir-se aflitivo aflito afobado afogado afogar as mágoas afogar-se em pouca água afogar-se afrontado afrontoso afundado afundação afundar-se agachado agarotado agarrado agarrar-se ao poder agastadiço agastado agastar-se agiota agir de má-fé agoirente agonia agoniado agonizante agonizar agonizante agoureiro agourento agravado agravar agre agredir agressividade agressivo agressor agreste agrilhado agrosseirado aguado aguçado alapado alarmante alarmismo alarmista alarve alarvice albino alcoolismo alcoolizado alegado aleijado aleivoso algemado alheado alhear-se alheio aliciado aliciar alienação mental alienado alienígena alquebrado altaneiro altivo aluado alucinação alucinado alucinar alumbrado alvoraçado amachucado amadorismo amaldiçoado amalucado amaneirado amarelado amarelo amargado amargo amargura amargurado amaricado amarrado amassado ambiguidade ambíguo ambiopia ambliopia ameaça ameaçador ameaçante amedrontado amedrontador amigo da onça amigo da pinga amigo dos copos amnésia amnésico amolecer amoral amoralidade amoralista amorfia amorfo amor-

tecido amotinado amputado amuado amuamento amuar amuo anafado analfabeto anão anárquico anarquista andante andar a apanhar bonés andar à deriva andar a dormir à sombra da bananeira andar a dormir andar a falar sozinho andar à nora andar a pensar na morte da bezerra andar a reboque de andar à sombra da bananeira andar ao deus dará andar aos bonés andar aos caídos andar aos papéis andar armado em carapau de corrida andar às aranhas andar às cavalitas de andar às cegas andar às turras com andar com a cabeça na lua andar com cara de caso andar com dor de corno andar com dor de cotovelo andar com macacos na cabeça andar com macaquinhos na cabeça andar com o coração nas mãos andar com os azeites andar de olhos vendados andar de trombas andar em pele e osso andar feito barata tonta andar feito com andar na boa vida andar na lua andar na má vida andar pelos cabelos andrajoso anémico anêmico anestesiado angustiado angustiante animal animosidade aniquilado anjinho anorético anorético anorexia nervosa anorexia anormal anormalidade anotia ansioso anti-republicano antidemocrata antidesportista antidesportivismo antiliberal antiliberalista antinatural antipatia antipático antipatriota antipatriotismo antipolítico antiquado antisocial apagado apagar-se apalermado apalhaçado apancado apanhado apanhar nas orelhas apanhar por tabela apaparicar aparolado apartado aparvalhado apatetado apatia apático apatriota apavorado apavorante apavorar-se apeado apedantado apedrejado apedrejar aperaltado aperreado aplicar golpes baixos a apoplético apoplético apoquentado apoquentador apreendido apreensivo apriorista aprisionado aprontar apropriar-se aproveitador aproveitar-se apunhalado apunhalar pelas costas apunhalar aquietador arbitrário arcaico ardil ardiloso árduo arfante árido arisco armar barraca armar-se aos cágados armar-se aos cucos armar-se em carapau de corrida armar-se em chico esperto armar-se em parvo armar-se arquejante arranhado arranjar lenha para se queimar arranjar sarna para se coçar arranjar um bico de obra arrasado arrastadeiro arrastado arrebentado arrebitado arredado arredio arreliado arrelizador arrenegado arrepender-se arrepiente arrepio arrevesado arrivista arrogante arrolado arruaceiro arrufado arrufo arruinado arruinamento arruinar-se arruinar arrumar as botas arteirice arteiro arteriosclerose artificial artificialidade artifício artificiosidade artificioso artimanha artritismo aselha aselhice asfixiante asilado asneira asneirento asno aspereza áspero asquerosidade asqueroso assalariado assaloiado assaltante assaltar assanhadiço assanhado assarapantado assassinar assassino assediado assediador assediar assédio sexual assédio assimétrico assoberbado assolado assolador assomadiço assombrado assustadiço assustado assustador assustar-se astenia asténico astúcia astucioso astuto atabalhoação atabalhoado atabalhoamento atacante atado atafulhado atamancado atarantação atarantado atarantar-se atarracado ateado atemorizado aterrado aterrador aterrorizado aterro-

rizador atirar areia para os olhos de atirar barro à parede atirar com as culpas a atirar poeira aos olhos de atoleimado atônito atordado atordoador atormentado atormentador atracado atraído atrapalhado atrapalhar-se atrasado mental atrasado atraso mental atraso atravessadiço atravessado atribulado atroador atrocidade atrofia atrofiado atrofiante atrofiar-se atrofiar atropelado atropelar atroz atulhado aturdido austero autismo autista autodestruição autodestrutivo autoritário autoritarismo avarento avareza avariado avariar avaro avassalador avassalamento avassalante averso avesso ávido aviltado aviltador aviltante avinagrado avinhado avulso azafamado azambrado azar azarado azarento azedar azedo azedume azeiteiro azelha azelhive azoinante azougado babaca babão baboseira babosice baboso baço para espelho bacoco bacoquice badalhoco badalhoquice bagaceiro bagunceiro baixeza baixote bajulador bajular balbuciante balofice balofo bambo banal banalidade banalizar banana bancarrota bandido banditismo banido banzado baralhado baralhar-se barato barbaridade bárbaro barrado barraqueiro barrigana barrigudo barroco barulhento baselga básico bastardo batatado bater a bota bater as botas bater com o nariz na porta bater no ceguinho bater no fundo batido batota batoteiro batotice beatice beato bêbado bebedeira bêbedo bebedor beberrão bebido bebum beíçudo belfo belicismo belicista bélico belicoso bem mandado bera besta bestificado bexiguento bibliomania bilioso biltre bimbo biqueiro bira birra birrento biruta bisbilhotar bisbilhoteiro bisbilhotice bisonho bizarria bizarro blasfemador blasfêmia blasfemo bloqueado bloquear boateiro bobo boboca boçal bocejar bochechudo boêmio boémio bojudo bolachudo bolorento bombeado borbulhento borguista borrachudo borrado boto botocudo bravateador brega brejeirice brejeiro brejeirote briga brigador brigante brigão briguento brilharete brincar bronco bronquice bruno brusco brutal brutalidade brutidão bruto bruxaria bucha bufismo bufo bulhão bulhento bulício buliçoso burguês burla burlado burlador burlão burlesco burlista burrice burro cabazada cabeça dura cabeçudo cabisbaixo caboclo cabotino cabrão cábula caçado cacete caceteiro cachaceiro cadastrado cadavérico cadaveroso cadente caduco cafona cagão cagar-se cagarola cagativo caguinchas caído caipira cair na armadilha cair na boca do lobo cair na ratoeira cair por terra cair calaceiro calão calar a boca calar o bico calar-se calcado calculismo calculista calhandreiro calhandrice calhorda calino caloteirismo caloteiro calotismo calúnia caluniado caluniador caluniar cambada cambaleante cambalear camelo canalha canalhive canceroso cancro candongueiro candonguice cangalheiro canhestro canibal canino cansado cansativo cantar bem mas não me alegra cantar de galo capcioso capenga capitoso capricho caprichoso caquético caquético caquexia cara de cu cara de pau caralho carapaça carbonizado carcomido carecer carecido careiro carência carenciado carente caricato carnavalesco carniceiro carrancudo

casmurrice casmurro casquilho cassete castigado castigador castração castrador catalepsia catano cataplexia catarro catatonia catinguento cativo caturra caturrice causador causticação causticante causticidade cáustico cavernícola cavernoso caviloso ceder cego cegueira cegueta celerado censurado cepo cercado chalaceiro chalado chamuscado chanfrado chantagista charlatanice charlatão chateado chatice chato chauvinismo chauvinista chavão cheio de nove horas cheirar a mofo chibante chicaneiro chico esperto chicoteador chifrudo chinesada chinesice chinfrim chinfrim chinfrineiro chocalheiro chocar chocarreiro chochice chocho choco choné choramingão choramingar choramingas chorão chorar sobre o leite derramado chorar choroso chulo chumbado chumbar chunga chupado chupador chupismo chupista cigano cilindrado cilíndrico cínico cioso cisma cismado cismático ciúme ciumento clamoroso clandestino claudicante claustrofobia cleptomania cleptomaniaco coagido coagir coartado cobarde cobardia cobardola cobiça cobiçoso cocainomania coercivo cogitabundo cogitação cogitativo coio coitado colaboracionista colapso colérico comandita combalido combalimento comédia comedor comer gato por lebre comer o pão que o diabo amassou comichento comichoso comilão comodismo comodista compactuar comparsa compelido compelir complexado complicado comprado comprar gato por lebre comprometedor compulsivo compungido comuna conceituoso condenado condenável condescendência condoído condoreiro conflituosidade conflituoso conformado conformar-se conformista confrangedor confranger-se confundido confundir alhos com bugalhos confundir-se confusão confuso congelado congestão congestionado conivente conluiado conspiração conspirador conspirativo conspurcar consternação consternado consternador consternar constipado constrangedor constranger constrangido constrangimento consumido consumir-se consumista contagiado contagioso contaminar contendedor contendente contentar-se com pouco contestado contido contingente contorcer-se contracto contradizer-se contraído contrair-se contranaturo contrariado contristação contristado contrito controlador controvérsia controverso controvertido contumaz contundência contundente contundido contuso convulsivo coquete corcovado corcunda cornudo cornuto correr a toque de caixa corrido corriqueiro corroer-se corroer corromper corrompido corrosivo corrupção corruptível corruptivo corrupto corruptor cortado cortar na casaca de coscuvilheiro covarde covardia coveiro coxo crápula crapuloso crasso cretinice cretinismo cretino criancice criminalidade criminoso crispação crispado crisar-se criticável crivado cromo cru crucificador cruel crueldade crueza cruzar os braços culpa culpado culpas culposo curto de vistas curvado curvar-se curvilíneo curvo cuspidor danado danificado danoso dar à língua dar à sola dar à soleta dar a volta ao miolo dar ao pé dar bandeira dar barraca dar cabo da vida a dar com a boca no trombone dar com a língua nos dentes dar com os burros na água dar com os pés a dar de frosques dar

graxa a dar o corpo ao manifesto dar o dito por não dito dar o golpe a dar o golpe do baú dar palmadinhas nas costas a dar pontapés na gramática dar tiros nos pés dar um pontapé no cu a dar um tiro no pé dar uma no cravo e outra na ferradura dar voltas no caixão deambulação deambulante deambular debandado débil debilidade debilitação debilitado debochado deboche decadência decadente decadentista decaído decepado decepção decepcionado decepcionante decepcionar decrepidez decrépito defeituoso defensiva defesso deficiente definhado definhamento deformação deformado defraudar as expectativas de defraudar defunto degenerado degradado degradante degredado degredo deicida deitar achas na fogueira deitar as unhas a deitar foguetes antes da festa deitar lenha na fogueira deitar poeira aos olhos a deitar poeira para os olhos de deitar tudo a perder deitar tudo para trás das costas deitar tudo por água abaixo deixar com as calças na mão deixar com as mãos a abanar deixar de rastos deixar muito a desejar deixar o barco à deriva deixar o país de tanga deixar tudo para depois deixar-se levar delambido deletério delinquência delinquente delirar delituoso demagogia demagogo demarcado demência demente demérito demitido démodé demolidor demoníaco demorado denegrido denegrir a sua imagem dengoso depauperado depenado dependurado deploração deplorável deportado deposto depravação depravado depravador depreciado depreciar depreciativo depressão maníaca depressão depressivo depressor deprimência deprimente deprimido derramado derreado derreamento derrota derrotado derrotismo derrotista derrubado desabonado desabrido desabrigado desabusado desacompanhado desacordado desacreditado desactualizado desadaptado desafecto desafeto desafinação desafinado desaforado desaforido desafortunado desagradado desagradável desagradecido desagrado desaire desairoso desajeitado desajustado desalentado desalento desalinhado desalinhamento desalmado desalojado desamparado desamparo desanimação desanimado desanimador desapaixonado desaparafusado desaparecer do mapa desaparecido desapegado desapiedado desapontado desapontador desapossado desapropriado desaprovador desaproveitado desarranjado desarrazoado desarrumado desarticulado desassinado desassossegado desassossego desastrado desastre desastoso desatento desatinado desatualizado desaustinado desavergonhado desavisado desbaratado desbaratador desbocado desbocamento desbotado desbragado descabelado descabelar-se descabido descadeirado descalabro descalço descamisado descarado descaramento descarnado descascado descautelado descer ao mais baixo nível descer nas sondagens descerrado desclassificado descomedido descomedimento desconcentração desconcentrado desconcertado desconexo desconfiado desconfortável desconhecedor desconsideração desconsolação desconsolado descontentamento descontente descontrolado descontrolo descoordenação descoordenado descorado descoroçoado descortês descortesia descuidado descuido desdém

desdenhar desdenhoso desdentado desdito desditoso deselegante desemparelhado desempregado desemprego desencalhe desencaminhado desencaminhador desencaminhamento desencontrado desengonçado desengraçado desenxabidez desenxabido desequilibrado desequilíbrio desertar desertor desesperado desesperança desesperançado desespere desestabilizador desfaltar desfalecido desfasado desfavorecido desfeito desfigurado desgarrado desgastado desgaste desgosto desgostoso desgovernação desgovernado desgoverno desgraça desgraçado desgracioso desgrenhado desidratação desiludido desinfeliz desinformação desinformado desinstruído desinteressado desinteressante desinteresse desistir deslambido deslavado desleal deslealdade desleixado desleixar-se desleixo desligado deslocado desmaiado desmanchado desmantelado desmazelado desmazelamento desmazelo desmedido desmemoriado desmentido desmesurado desmiolado desmontado desmoralização desmoralizado desmoralizador desmorrar-se desmotivação desmotivado desmotivador desnaturado desnorteado desnorteamento desnutrição desobediência desobediente desocupado desolação desolado desolador desolhado desonestidade desonesto desonra desonrado desonrar desonroso desordeiro desordem desordenado desorganização desorganizado desorientação desorientado despassarado despeitado despeitar despeito despeitorado despejado despenteado despercebido desperdiçado desperdiçador despesista despirocado despistado desplante despregado despreocupação desprevenção desprevenido desprezado desprezável desprezível desprezo desprimoroso desproporcionado despropositado desprotegido despuador despudorado desqualificado desregrado desregulado desremediado desrespeitado desrespeitador desrespeitar desrespeito desrespeitoso destabilizador destampado destapado destemperado desterrado destoante destrambelhado destrambelhamento destratado destravado destroçado destroço destruição destruído destruidor destrutivo destrutor desumanidade desumano desunido desusado desvairado desvairamento desvalido desvanecer desvantagem desvariado desvario desventurado desventuroso desviar dinheiro desvirtuado desvitalizado deteriorado deteriorar detestado detestável detido detrator deturpado deturpar devaneador devassado devassidão devasso devastado devastador dever comer farinha maizena dever dobrar a língua dever ter vergonha na cara devoto de Baco diabolismo difamação difamado difamador difamar difícil difuso digressionista digressivo dilacerado diminuído díscolo discriminação discriminado discriminador discriminar disfarçado disfonía disforme disformidade dislexia disléxico disparatado disparatar disparate dispensável dispersivo disperso displicência displicente dissidente dissimulação dissimulado dissimular dissipado dissipador dissoluto dissolvido distante distorcer distorcido distraído ditador divagação divagador divagante divagar dizer cobras e lagartos de dizimado doente doentio dogmático dogmatista doidivasas doido de todo doido varrido doido dolente dolo-

rido doloso dominado dopado dopagem dorido dormência dormente dorminhoco dormir
doudo draconiano dramático dramatismo dramatizar drástico droga drogado dromoma-
nia dromomaníaco dubiedade dúvida dubitativo dificuldades dúplice duplicidade duro da
mioleira duro de ouvido duro duvidoso edipiano efeminado egoísmo egoísta egotista ele-
gíaco eleiçoeiro elitista emaciado emaranhado embaçado embaciado embaraçado embaraço
embaraçoso embasbacado embatucado embirração embirrente embirrento embotado em-
briagado embrulhado embruxado embuçado embuchado emburrado embuste embusteiro
empáfio empalidecido empalmado empanturrado empapuçado empedernido empederni-
mento emperrado emperramento empertigado empobrecido empoeirado empolado em-
porcalhado emproado emulador êmulo encabulado encalacração encalacrado encalhado
encalhamento encamado encanado encarcerado encarniçado encavacado enciumado en-
clausurado encoberto encobrir encolerizado encolerizar-se encolher os ombros encolhido
encostar-se a encovado encruado encurralado encurralamento encurvado endemoniado en-
demoninhado endiabrado endividado endividamento endividar-se endividar endoidecer
endurecido enervação enervado enervamento enervante enervar-se enfadado enfadamento
enfadar-se enfadonho enfardador enfartado enfartamento enfatiado enfatiamento enfas-
tante enfatuado enfeitado enfermiço enfermidade enfezado enfezamento enfiado enfiar o
barrete enforçar-se enfraquecer enfraquecido enfrascado enfronhado enfunado enfurecer-se
enfurecido enfurecimento engaiolado engaiolamento enganado enganador enganar-se enga-
nar engano enganoso engarrafado engasgado engasgamento engatado engodado engolfado
engolido engolir em seco engonha engrupido enigmático enjeitado enjoado enjoar enjoo
enlameado enleado enlevado enlouquecer enlouquecido enlouquecimento enlutado enojado
enojar enraivecido enraivecimento enrascada enrascado enredado enregelado enresinado
enrijecido enrodilhador enrolado enrugado ensacado ensanguentado ensanguentado en-
simesmado ensimesmamento ensurdecido entalado entalar-se entaramelado entediado
entediante enterrado enterrar a cabeça na areia enterrar-se entontecido entontecimento
entornado entorpecer entorpecido entorpecimento entortado entregador entregar a alma
ao criador entregar a alma ao diabo entregar o ouro ao bandido entrevação entrevado
entristecer entristecido entristecimento entubado entulhado entumescimento envaidecido
envelhecido envenenador envenenamento envenenar envergonhado envergonhar-se enver-
gonhar enviesado envinagrado enxerido enxofrado enxoframento enxovalhado epigramá-
tico epilético equivocador equivocarse equívoco equívoco ermo erotomania erotomaníaco
errado errante errar errático erro esbaforido esbanjado esbanjador esbanjamento esban-
jar dinheiro esbatido esborratado esburacado escabelado escabrosidade escabroso escal-
dado escamado escamoso escamotear escancarado escandalizar escandaloso escangalhado

escanifrado escanzelado escapadiço escapo escarnecedor escarnecido escarninho escárnio escarrapachado escavacado escaveirado esclerosado esclerose esconder a verdade esconder-se esconder escondido esconso escoriado escorraçado escorregadio escorregar escrachado escravizado esculachado escuro escusado escuso esdrúxulo esfaimado esfalfado esfarrapado esférico esfíngico esfomeado esgalgado esganação esganado esganiçado esgazeado esgazeamento esgotado esgotamento esgotante esgrouviado esguedelhado esmagado esmilhafrado esmorecer esmorecido esmorecimento espalhafato espalhafatoso espalhanço espampanante espancado esparramado espatifado espaventoso espavorido especado espertalhão espetaculoso espetado espetar-se ao comprido espezinhado espinhento espinhoso espremido espúrio esquálido esquecediço esquecido esquelético esquentado esquinado esquisitice esquisito esquivo esquizofrenia esquizofrénico esquizoidia estabanado estacado estafado estagnado estagnar estalado estalar o verniz estampado estancado estanhado estapafúrdio estar a anos luz de estar à beira do precipício estar à deriva estar a dormir à sombra da bananeira estar a dormir estar a falar para as paredes estar a falar sozinho estar a jeito estar à nora estar a pensar na morte da bezerra estar à rasca estar à rasquinha estar a secar estar à sombra da bananeira estar ao deus dará estar armado em carapau de corrida estar às aranhas estar às cegas estar às turras com estar com a cabeça na lua estar com a corda ao pescoço estar com a corda na garganta estar com a corda no pescoço estar com a lágrima no canto do olho estar com cara de caso estar com dor de corno estar com dor de cotovelo estar com macacos na cabeça estar com macaquinhos na cabeça estar com o coração nas mãos estar com os azeites estar com os pés para a cova estar com um aperto no coração estar de cabeça quente estar de mãos a abanar estar de mãos atadas estar de mãos e pés atados estar de molho estar de olhos vendados estar de pés atados estar de tanga estar de trombas estar em baixo de forma estar em baixo estar em pedaços estar em pele e osso estar enterrado até às orelhas estar feito ao bife estar feito barata tonta estar feito com estar fora de si estar na lua estar na retranca estar nas mãos de estar num beco sem saída estar pelos cabelos estar por um fio estar sempre a bater na mesma tecla estar-se nas tintas estarola estarolice estarrecido estatelado estatelar-se estático estavanado estendido estereotipado estéril esterilidade esticado esticar o pênfil estigmatizado estiloso estirado estoirado estoiro estourado estouvado estrábico estragado estrangeirado estrangulado estranho estratagema estremunhado estridente estroina estropiado estruturado estulto estupefaciente estupidez estúpido estuporado estuprado esvair-se em sangue esvaizado etilizado eunuco evasão evasivo evitar exacerbado exagerado exagerar exalado exaltado exaltar-se exangue exasperação exasperado exasperar-se exasperar exaustão exausto exceder-se excessividade excessivo excluído excomungado execração execrado execrável

executado exibicionismo exibicionista exibido exhibir-se exilado eximido exotérico expatriado expelido experimentar dificuldades explorador explorar explosivo expropriado expulso extenuação extenuado extenuante extravagante extraviado extremista façanhudo faccioso facínora facinoroso fadiga fadigado falácia falacioso falar de barriga cheia falar demais falar para as paredes falar para o boneco falar sem convicção falência falhado falhanço falhar falho falibilidade falido falível falsear falseta falsidade falso faltar à palavra faltar ao compromisso faltar ao respeito falta faltoso famélico faminto fanfarrão fanho fanhoso fantasiar fantasista fareleiro farisaico fariseu farrista farronqueiro farsa farsante farsista farsola fascista fastidioso fastio fatal fatalista fatigado fatigante fátuo favorecer fazer à custa de fazer a folha a fazer asneira fazer beicinho fazer biquinho fazer caretas fazer faísca fazer figura de urso fazer figura triste fazer filmes fazer gato e sapato de fazer horas fazer má figura fazer orelhas moucas fazer ouvidos de mercador fazer uma rica parelha fazer uma tempestade num copo de água fazer vista grossa fazer-se de vítima febril fechado fechar os olhos fechar-se em copas fedor fedorento feio feiticista feitiço felino femeeiro ferido ferino fero feroz ferrado ferrenho férreo ferver em pouca água fescenino fetichista ficar a apanhar bonés ficar a chupar no dedo ficar à deriva ficar a dormir à sombra da bananeira ficar a falar para as paredes ficar a falar sozinho ficar a jeito ficar à nora ficar a pensar na morte da bezerra ficar a secar ficar à sombra da bananeira ficar à sombra de ficar a ver navios ficar ao deus dará ficar aos caídos ficar às aranhas ficar às cegas ficar com a cabeça em água ficar com a cabeça na lua ficar com as calças na mão ficar com cara de caso ficar com dor de corno ficar com dor de cotovelo ficar com macacos na cabeça ficar com macaquinhos na cabeça ficar com o coração nas mãos ficar com os louros ficar de mãos a abanar ficar de mãos atadas ficar de mãos e pés atados ficar de molho ficar de pés atados ficar de tanga ficar de trombas ficar em pedaços ficar em último ficar enterrado até às orelhas ficar feito barata tonta ficar fora de si ficar mal visto ficar na lua ficar nas mãos de ficar num beco sem saída ficar para tio ficar pelos cabelos ficar reduzido a nada ficar reduzido a zero ficar sem argumentos fictício filho da puta finado fincado fingido fingidor fingimento fingir fígado fiteiro fitinhas fiúza flagelado flagelante flatulência flatulento flébil fodido fofoqueiro fofoqueice foleirice foleiro folgado fome foragido forasteiro forçado fornicador forreta fosco fracalhote fracassado fracassar fracasso fraco fracote frágil fragilidade fragilizado fragmentado franzino fraqueza fraturado fraude fraudulento frenético frieza frigidez frígido frio frito frivolidade frívolo frouxidão frouxo fruste frustração frustrado frustrar fugidio fugido fugir à questão fugir à responsabilidade fugir a sete pés fugir ao fisco fugir às perguntas fugir às responsabilidades fugir com o rabo à seringa fugir fugitivo fujão fulminante fulo fundamentalista funesto furi-

bundo furioso furtar furtivo fútil futilidade gabar-se gabarola gabiru gadelhudo gafo gagá gago gaguez gaiteiro galdério gamar ganância gandulagem gandulo gangrena ganir ganizado garganeiro garotice garoto gaseado gastador gastar mundos e fundos gasto gatuno gaudério gelado gélido genioso gentinha gentio gingão glutão goliardo golpeado gorado gordanchudo gordo gorducho gordurento gorduroso gozador gozo grafomania grafomaniaco grazina gringo gritante grosseirão grosseirismo grosseiro grosseria grosso grotesco grudento guerra guloso handicap hediondez hediondo hemiplegia herege heresia herético hermético herodiano heroinomania hesitante heterogêneo heterogêneo hiperactivo hiperactivo hiperconservador hiperconservadorismo hiperocrítico hipersensibilidade hipersensível hipnotização hipnotizado hipnotizar hipnotizável hipocondria hipocondríaco hipocrisia hirsuto histeria histérico homicida homicídio homiziado horrendo horribilidade horrífico horripilação horripilante horrível horror horrorizar horroroso hostil hostilidade humilhado humilhar-se idiota idiotia idiotice ignaro ignomínia ignominioso ignorado ignorante ignoto ilegal ilegítimo ilegível iletrado iliberal iliteracia iliterato ilógico ilogismo iludido iludir ilusão ilusório imaturidade imaturo imbecil imbecilidade imerecido imisericordioso imitador imitar imitativo imobilizado imoderação imoderado imodéstia imodesto imoral imoralidade imorigerado imóvel impaciente impante impatriota impecunioso impelido impenetrável impenitente imperceptível impercetível imperdoável imperfeição imperfeito imperial imperialista imperiosidade imperioso imperito impermanência impermanente imperscrutável impertinência impertinente impessoal impessoalidade impetuosidade impetuoso impiedade impiedoso ímpio implacável implicado implicante implicativo implorar imponderação imponderado impontual impontualidade impopular impopularidade importunado importuno impossível imposto impostor impostura impotência impotente imprecisão impreciso impregnado impreparado impressionável imprestabilidade imprestável imprevidência imprevidente improcedência improcedente improdutora improdutividade improdutivo imprudência imprudente impudente impudícia impudico impugnado impulsivo impuro imputado imundície imundo inábil inabilidade inabordável inaceitável inacessível inactividade inactivo inadaptado inadequado inadmissível inamovível inanimação inanimado inânime inaplicado inapresentável inapto inatividade inativo inautenticidade inautêntico incapacidade incapaz incaracterístico incauto incendiado incendiário incensador incerteza incerto inchado incipiente incivil incivilidade incivilizado inclemência inclemente incoerência incoerente incógnito incomodado incomodativo incomodidade incômodo incômodo incompetência incompetente incompleto incompreensibilidade incompreensível inconcebível inconciliável inconcluso inconfessável inconfidente incongruência incongruente inconsciência inconsciente inconsequente inconsequente inconsiderado inconsistência in-

consistente inconsolado inconsolável inconstante incontinência incontinente incontrariável
incontrolado inconveniência inconveniente inconversável incorreção incorrecção incorrecto
incorreto incorrigível inculto incultura incumpridor incurável indecência indecente indeci-
frável indeciso indecoro indecorosidade indecoroso indefenso indeferido indefeso indefinido
indelicateza indelicado indesculpável indesejado indesejável indeterminado indiciado in-
digente indigesto indigitável indignado indignidade indigno indirecto indireto indisciplina
indisciplinado indiscreto indiscrição indisposição indisposto indistinto inditoso individua-
lista indócil indocilidade indolência indolente indominável indômito induzir em erro ine-
briado inebriante inefável ineficácia ineficaz ineficiência ineficiente inelegível inenarrável
inépcia inepto inércia inerme inerte inescrutável inexactidão inexacto inexactidão inexato
inexorável inexperiência inexperiente inexperto inexplicável inexplicado inexpressivo inex-
primível infamado infamante infame infâmia infantil infantilidade infectado infecundo
infelicidade infeliz infernal infértil infetado infidelidade infiel inflado inflamado inflexibi-
lidade inflexível influenciável infoexcluído informe infortunado infortúnio infrene infringir
infrutífero infundado ingênuo ingratitude ingrato inibido inibir-se inidoneidade inidóneo
inimigo ininteligibilidade ininteligível iniquidade iníquo injúria injuriado injuriante inju-
riar injurioso injustiça injusto inoportuno inóspito inqualificação inqualificado inqualifi-
cável inquebrantável inquietação inquietante inquieto insaciado insanidade insano insa-
tisfatório insatisfeito insegurança inseguro insensatez insensato insensibilidade insensível
insidioso insignificante insinceridade insincero insinuação insinuador insinuativo insinu-
oso insipidez insípido insipiência insipiente insociabilidade insociável insofrido insolência
insolente insólito insondável insonso insosso instabilidade instável instigado insubmisso
insubordinação insubordinado insuficiente insulsez insulso insultado insultar insulto in-
sultuoso insuportável insurgente insurrecto insurreto insusceptível insuscetível integrista
intelectualóide intemperado intemperamento intempestividade intempestivo interceptado
intercetado interdito interesseiro internado interrogatório intimidação intimidado intimi-
dador intimidante intimidativo intimidatório intolerância intolerante intolerável intragável
intranquilidade intranquilo intransigência intransigente intratabilidade intratável intriga
intrigante intriguista intrincado intrometer-se intrometido introvertido intruso inumani-
dade inumano inútil inutilidade invadir a privacidade de invadir inválido invasivo invasor
invectivo inveja invejar invejoso invertido invetivo inviável inviesado invisível ir ao chão ir
ao tapete ir aos arames ir às paredes ir dentro ir desta para melhor ir para o outro mundo
ir pelo caminho errado ir pelos ares ir-se abaixo ir-se ao ar ir-se aos ares iracúndia iracundo
irado irascibilidade irascível irônico irónico irracional irracionalidade irrealismo irrealista
irregular irregularidade irrelevante irreligioso irrequieto irresoluto irrespeito irrespeitoso

irresponsabilidade irresponsável irreverência irreverencioso irreverente irridente irrisório irritabilidade irritação irritadiço irritado irritante irritar-se irritativo irritável isolamento jactância jactante jalofo javardice javardo jeremias jocosidade jogar ao faz de conta judicativo jururu kafkiano lábia labrego labreguice labroste lacónico lacónio lacrimoso ladino ladrão ladroagem laical laico lambão lambareiro lambeiro lamber as botas a lambuzado lamecha lamechar lamechice lamentação lamentar-se lamentável lamentoso lampeiro lamúria lamuriante lamuriar-se lamuriento lamurioso lancinante langoroso languescente lânguido laparoto lapuz lástima lastimar-se lastimável lastimoso lata latente laurear a pevide lavar roupa suja laxista lazarento lãzudo leigo lelo lentidão lento leproso lerdice lerdo lesado lesionado lesivo leso letal letargia letárgico levado da breca levado das maleitas levado do diabo levantado levar com a porta na cara levar com um balde de água fria levar na corneta levar nas orelhas de levar nas orelhas levar porrada levar um baile de levar um baile levar um banho de água fria levar um pontapé no cu levar um tiro levar uma abada levar uma cabazada levar uma coça levar uma malha levar uma surra levar uma tarefa levandade leviano libertário libertinagem libertino libidinoso licencioso lícito ligar a cassete liliputiano limitado limpar o sarampo a limpar o sebo a linfatismo lingrinhas linguareiro linguarudo lívido lixado lixar-se lixar logrativo lorpa louco loucura louvaminheiro ludibriado ludibriante ludibriar ludibrioso lugar-comum lúgubre lunático lusco lusófono luto lutuoso luxúria má apresentação má cara má disposição má educação má fama má formação má fortuna má-fé má-governança má-prestação macabro macaco maçado maçador macambuzice macambúzio macanço macareno macavenco machismo machista machucado macilento maçado maculado madraço mafeitoria máfia mafioso magalomania magano magoar magrelo magricela mal-afamado mal-afeiçoado mal-afortunado mal-agradecido mal-amado mal-cheiroso mal-comportado mal-educado mal-encarado mal-feito mal-humorado mal-intencionado mal-pago mal-vestido malabarismo malabarista malandreco malandro malcheiroso malcriado maldade maldição maldisposto maldito maldizente maldoso maledicência maledicente maleducado maléfico malencarado malevolência malevolente malévol malfadado malfazejo malfeito malfeitor malfeitoria malformação malformado malhado malícia malicioso maligno malnutrição malogrado malquisto malsão malsucedido maltrapilho maltratado maltratar maluco maluquice malvadez malvado malvisto mamalhudo mamar nas tetas de mamudo manchado manchar a imagem de manco mancomunado mandão mandar bocas mandar calar mandrião maneirista maneta manha manhoso mania maníaco-depressivo maníaco manietado manipulação manipulado manipulador manipular manipulável manobrado manobrar manteigueiro maquiavélico maquiavelismo maquiavelista maquinado marau marginal marginalidade marginalizado marginalizar marialva

maricão maricas mariola marioneta mariquice mariquinhas marmanjo maroto marralheiro
marrano marrão marreco mascarado masochista masoquismo masoquista massacre mas-
sado matador matar o vício materialismo materialista matias matraquear matreiro matuto
mau agouro mau agrado mau aluno mau cheiro mau comportamento mau desempenho
mau exemplo mau feitio mau gênio mau humor mau pagador mau maus hábitos mavór-
cio mazombice mazombo mazorro medíocre mediocridade meditabundo medo medonho
medricas medroso mefistofélico megalomania megalomaníaco megalômano megalómano
meio-alterado meio-apalermado meio-doido meio-morto melado melancólico melindrado
melindrar-se melindrar melindroso melodramático melómano mendacidade mendaz men-
dicante menosprezante menosprezar menosprezável menosprezo mentecapto mentir men-
tira mentiroso compulsivo mentiroso inveterado mentiroso mercantil mercantilista mer-
cenário merda merdoso merecer o que tem merencório mergulhado mérito mesquinhez
mesquinho metafísico metedição meter a carroça à frente dos bois meter a foice em seara
alheia meter a pata na poça meter a viola no saco meter água meter medo meter o carro à
frente dos bois meter o nariz onde não é chamado meter o pé na argola meter o pé na poça
meter o rabinho entre as pernas meter o rabo entre as pernas meter os cornos a meter
os dedos pelos olhos de meter os dedos pelos olhos dentro meter os patins a meter os pés
pelas mãos meter uma cunha meter-se às cavalitas de meter-se na alheta meter-se na boca
do lobo meter-se numa embrulhada meter-se onde não é chamado metido mexeriqueiro
migalheiro mijão mimado mimalho minado mingua minúsculo míope mirabolante miro-
lho mirração mirrado misantropo miserabilismo miserabilista miserando miserável mísero
misturar alhos com bugalhos mitomania mitómano mocado modorrento modorro moedor
mofino moído moinante mole molenga molengão molenguiça moleque molestado molestar
molieresco mongolismo mongolóide mono monocórdico monomania monomaníaco mono-
polista monopolizado monopolizador monotonia monótono monstruosidade monstruoso
moquenco moralista morbidez mórbido morboso mordacidade mordaz mordente morder
a língua morder o isco morfinomania morfinómano moribundo morno moroso morrer de
pé morrer na praia morte mortício mortífero mortificado morto mostrar a verdadeira face
mostrar dificuldades mouqueira mouquice mudar as regras a meio do jogo mudar de ideias
como quem muda de camisa mudez mudo mula mulherengo mundano murcho murmurante
mutilado nabice nabo nanico não acertar uma não chegar aos calcanhares de não chegar
aos pés de não chegar às solas dos sapatos de não dar uma para a caixa não dizer ai nem
ui não dizer coisa com coisa não dizer nada de jeito não estar à altura de não estar em
si não fazer a ponta de um corno não gostar de ouvir as verdades não interessar a nin-
guém não mexer uma palha não olhar a meios para atingir os fins não olhar a meios não

perceber patavina não saber o que dizer não saber o que quer não saber onde se meteu não saber patavina não sair da cepa torta não se aprender nada com não se mexer não ser carne nem peixe não ser levado a sério não ser peixe nem carne não ter argumentos não ter nada na cabeça não ter nada nos miolos não ter onde cair morto não ter pedalada não ter razão não ter um pingo de vergonha não ter um tostão não ter vergonha na cara não ter vergonha nenhuma não valer nada não ver um boi à frente dos olhos não ver um boi não ver um palmo à frente do nariz narcisista narcótico narcotizado narigudo nauseado nauseante nazi nazista nebuloso necessitado necrófago necrófilo nédio nefando nefasto nefelibata negaceiro negatividade negativo negligência negligenciar negligente negregado negróide nervos nervosismo nervoso néscio neurasténico neurose neurótico ninheiro ni-
quento nocivo nódoa nojento nojo noveleiro novidadeiro nulo número de circo obcecação obcecado obcecante obeso oblíquo obrigado obscenidade obsceno obscurecer obscurecido obscuridade obscuro obsessão obsessivo obsoleto obstinado obstrucionista obstruir obtusidade obtuso ociosidade ocioso oco ocultar a verdade ocultar oculto odiado odiável odiento ódio odioso ofegante ofender ofendido ofensa ofensivo oferecido ofuscado oleoso olheirento oligofrenia oligofrénico omissio omitido omitir ondulante onerado onzeneiro opaco opiniático opiomania opiómano oponente oportunismo oportunista opressão opressivo opresso opressor oprimente oprimido opulento ordinarice ordinário orelhudo órfão orgulhoso os-
sido ostensão ostensivo ostentação ostentativo ostentoso ouvir das boas ouvir poucas mas boas pacovice pacóvio padecedor padecente pagar caro pagar favores a pagar um favor a palavroso paleio palerma palermice palhaço palidez pálido panal panasca pancadaria pançudo panema pantafaçudo pântano pantemineiro papaguear papalvo papão papista parado paradoxal paradoxo paralisado paralítico paranóia paranóico parasita parasito pa-
rido parolice parolo parracho parrudo partidarista partido partir o coração a parvalhão parvo parvoíce parvónio pascácio pasmado paspalhão paspalho passadista passado passar a batata quente a passar à história passar a perna a passar cheques em branco passar desta para melhor passar dos limites passar fome passar por dificuldades passeiro passivo pastelão pastrano pataqueiro patarata patau patego pateta patetice patético patibular pa-
tifaria patife patinar patola patrioteiro patudo patusco pau de cabeleira pavonear-se pavor pavoroso pecado pecador pecar pechoso peco pedante pedantice pedantismo pedinte pe-
drado pegajoso peganhento peguilhento peitado pelintra penalizado pencudo pendurado pendurar as botas peneirento penetra penitenciar-se penoso pequenez perdedor perder a cabeça perder a compostura perder a postura perder a razão perder as estribeiras per-
der o juízo perder o norte perder o pio perder os sentidos perder terreno perder perdido perdulário perecedor perecível perfídia pérfido perigo perigosidade perigoso permissivo

perneta pernicioso pernóstico perplexo perro perseguir persuadido pertinácia pertinaz perturbação perturbações de carácter perturbações de caráter perturbado perturbador perturbante perversidade perverso pervertedor perverter pervertidade pervertido pesadão pesaroso pesporrência pesporrente pessimismo pessimista péssimo pétreo petrificado petrificante petulância petulante piada picado picante picareta falante pícaro picuinhas piegas pieguice píffio pigmeu pimpão pindérico pinguço pinoca pintalegrete piolhoso pior piorar pipi piração pirado piranga pirangueiro pirar pirata piromania piromaníaco pirómano pirosice piroso pisado pisar na bola pisar o risco pisco pitónico pitosga plangência plangente plebeu pobre coitado pobre de espírito pobre e mal agradecido pobre pobretana pobretão pobrete pobreza de espírito pobreza podre poleiro polémica polémico polemista politicante politiqueiro politiquice poltrão poluir pontudo populismo populista porcalhão porcaria porcino porco porfiado pornografia pornográfico possessão possessividade possessivo possesso possidónio possuído postiço precariedade precipitado precipitar-se precisado precisar de comer farinha maizena precoce preconceito preconceituoso pregar aos peixes preguiça preguiçoso prejudicado prejudicar prejudicial premeditado preocupado prepotência prepotente presbíope presbiopia presbitismo presidiário preso presumido presumível presunção presunçoso pretensiosismo pretensioso pretenso primitivo prisão prisioneiro problema problemático processado profanador profano profuso proibitivo prolixo prometer mundos e fundos promiscuidade promíscuo pronóstico prosaico proscrito prosista prostituído prostituir-se prostituto prostrado protervo protestação protesto provincianismo provinciano provocação provocador provocante provocativo provocatório provocável pseudo-intelectual psicopata psicopatia psicose psicótico pueril pulha pulhice pulverizado punido puritano púrrio pusilânime puto putrefato pútrido puxadinho quadrado quadrúpede quebrado quebrantado quebrantar quebrar a promessa eleitoral quebrar a promessa quebrar as regras do jogo quebrar as regras quebrar queimado queimar tempo queimar-se queixinhas queixoso queixudo queque querer poleiro querer tacho questionável quezilento quezília quixotesco rabelaisiano rabeta rabicho rabino rabudo rabugento rabugice rabulista racista radicalista raiva raivento raivoso ralado ralador ralhão ralheta rameloso rancor rancores rancoroso rançoso ranheta ranhosice ranhoso ranzinza rapioqueiro raposeiro raposino raquíptico raquitismo rasca rasgado rasteiro rastejante reaccionário reacionário reactivo reativo rebaixado rebaixar-se rebarbativo rebatido rebelado rebelde rebuscado recalcado recalcamiento recalcitrante receoso rechaçado rechonchudo recluso recriminador recusado redondo reducionista redundância redundante redutor reduzido refalsado refilão refractário refratário refugiado refugiar-se refutado regatão regateador regateiro regelado reincidente rejeitado relambório relapso relaxado relel remelado remeloso remendado re-

mendão remendeiro remisso render-se renegado repelente repelido repentinoso repetente repetitivo replicador repreensível represso repressão repressivo repressor reprimido reprovação reprovado reprovador reprovável reptante réptil repugnância repugnante repulsão repulsivo repulso requebrado requebro resignado resmungão resmunguice respingão respondão respondedor ressabiado ressentido ressequido restringido retardado reticente retido retorcido retórico retorto retracto retraído retrain-se retrete retrocesso retrogradação retrógrado retumbante réu revel revelho revoltado revoltante revoltoso revolucionado rezina rezingão ribaldeiro ribaldo ridicularização ridicularizador ridicularizante ridicularizar-se ridiculez ridículo ridículo rigorista risível rispidez ríspido rodado roedor roer-se rolê roliço rompido roncador roncante ronção ronceiro rosnar roto roubado roubar rouco roufenho rouquidão rude rudeza rudimentar ruído ruidoso ruim ruindade ruinoso russóforo rústico sabichão sabotar sacado sacana sacrifício sacrilégio sacrílego sacrista sacudir a água do capote sádico sadismo sadista safado sáfaro sáfio sair de mãos a abanar sair melhor do que a encomenda sair o tiro pela culatra sair-se mal saloioce saloio saltão saltar para o poleiro saltitante sandeu sangrento sanguessuga sanguinário sanguinário sanguinolência sanguinolento sanha sanhudo sapeca sapudo saquear sarcástico sardanapalesco sardónico sarna sarnento sarnoso satânico satanismo satírico saturado saturante sebentice sebo sebo seco sectário sedentário sedente sedento sedicioso segregado seguir o caminho errado selvagem selvajaria selvático sem-carácter sem-caráter sem-sabor sem-sal semibárbaro semicivilizado semiconsciente semilouco semimorto seminu semivivo senil senilidade sensaborão sensacionalista sentenciado sentencioso sentir dor de corno sentir dor de cotovelo sequestração sequestrar sequestrável sequestro sequioso ser culpa de ser do contra ser do mesmo saco ser farinha do mesmo saco ser mau como as cobras ser melhor do que a encomenda ser um zero à esquerda ser uma bela prenda ser uma cópia mal feita ser uma cópia mal produzida ser vítima de si próprio serelepe sertanejo servil servilismo servo severo sevo showoff sibilino sifilítico simplista simplório simulado sinistro sinuoso sudo snob snobe snobismo sobrealimentado sobrecarregado sobressaltado sobressalto soez sofista sofedor sôfrego sofrer sofrido sofrimento solerte solidão solitário soluçante sombrio sonâmbulo songa-monga songamonga sonolência sonolento sonsice sonso sopeiro sordidez sórdido sorna sornateiro sorrelfo sorriso amarelo sorumbático soturno sovado sovina sovinice stress stressado subjectivo subjetivo subjugação subjugado subjugador subjulgante submergido submerso submetido submissão submisso subordinado subordinante subornado subornador subornar subornável suborno subserviência subserviente subversão subversivo subversor sucumbir súdito sufocado sufocador sufocamento sufocante sugar suggestionável suicida suicidar-se suicídio suíno sujidade sujo sumição sumido superdefensivo

superficial superficialidade supérfluo superlotado supersensível supersticioso suplantado
suplicar suposto suprimido surdez surdo surrado surreal surrealista surro susceptibilidade
susceptível suscetibilidade suscetível suspeito suspicaz susto tabagista tablóide tabu ta-
canhez tacanho tachado tácito taciturnidade taciturno taful tagarela tamanco tamaninho
tanso tapado tara tarado sexual tarado tarasco tardíloquo tardio tardo tareco tarefa ta-
rimbeiro tarouco tarouquice tartamudo tartufo tato teatral teatralidade teatralizar tédio
tedioso teimoso telhudo temer temerário temeroso temido temível temoroso temperamen-
tal tempestivo tempestuoso temudo tendenciosidade tendencioso tenebrosidade tenebroso
tenro tensão tenso ter a cabeça em água ter a consciência pesada ter a corda ao pescoço
ter a lágrima ao canto do olho ter as costas quentes ter as horas contadas ter culpas no
cartório ter de comer farinha maizena ter dor de corno ter dor de cotovelo ter falta de chá
ter má consciência ter macacos na cabeça ter macacos no sótão ter macaquinhos na cabeça
ter macaquinhos no sótão ter mais olhos que barriga ter mau perder ter memória curta
ter o nariz empinado ter o pavio curto ter o rei na barriga ter os dias contados ter sentido
de justiça ter sentido de oportunidade ter sentido de responsabilidade ter um aperto no
coração ter uma grande cara de pau ter uma grande lata terreno terrificante terrífico
terrível terror terrorífico terrorismo terrorista teso testudo tétrico tetro tigre tigrino tí-
mido timorato tinhosice tihoso tirania tirânico tiranizar tirano tirar a água do capote
tirar nabos da púcara tirar o cavaleiro da chuva tirar o cavalo da chuva tirar partido da
situação tirar proveito eleitoral tiririca tiritante tísico titubeante tocado tocador toldado
toleirão tolhido tolhimento tolo tomar um banho de água fria tombado tontice tontinho
tonto tontura torcer o nariz torcido tormento tormentoso torpe torto tortuosidade tortuoso
tortura torturado torturador torturante torvo tosco totó tóxico toxicodependência toxico-
dependente toxicomania toxicómano traficante trafulha trafulhice tragédia trágico traição
traíçoeiro traído traidor trair trama tramado trampa trampolineiro tramposso tranga-
lhadas transgredir transido transigente transtornado transtorno transviado trapaceiro
trapacento trapalhada trapalhão trapalhice traquina tratante trauliteiro trauma travesso
travessura travestido trazer água no bico trazer nas palminhas das mãos trazer o diabo
no corpo trazer o diabo no ventre trazer o rei na barriga tremebundo tremelicas tremente
tremar tremido trêmulo trémulo tresloucado tresmalhado treta trevo trincado triste tris-
teza tristonho triturado trivial troca-tintas trocar alhos por bogalhos trocar os pés pelas
mãos trocar-se todo trocista troglodita troliteiro trombudo trôpego truanesco trucidado
trucidante truculência truculento truque de magia truque tumulto tumultuado
tumultuário tumultuoso turbador turbulência turbulento turrão ultrajado ultrajante ul-
trajar ultraje ultrapassado untuosidade untuoso urinado usar usurário usurpador usurpar

utopia utópico utopista uxoricida vacilação vacilar vadiação vadiagem vadio vagabundagem vagabundo vagamundo vagaroso vago vaguear vagueza vaiado vaidoso valetudinário vandálico vandalismo vândalo vangloriar-se vão varado varrido vassalo vazio vedado velado velhacaria velhaco velho e relho velho velhote venal vencido vencível vendedor vender banha da cobra vender gato por lebre vender-se vendido veneno venenoso ventrudo ver a vida a andar para trás ver a vidinha a andar para trás vergado vergonha vergonhoso vermelhusco verrinoso vesgo vexação vexado viciado vicioso videirinho vigarice vigarista vígaro vil vilão vinagrento vingança vingativo vinolência vinolento violado violência doméstica violentado violento viperino vira-casacas virar a cabeça a virar a casaca virar as costas a virar o bico ao prego virulento visceral viscoso vítima vitimização vociferação vociferador vociferante voltar à cepa torta voltar as costas a voltar com a palavra atrás voluntarioso volúvel vomitado vomitar-se vômito vorace voraz vulgar vulgaridade vulgarizado vulgarizador vulnerabilidade vulnerável xenofobia xenófobo xexé xingar zaino zanga zangado zarolho zelote zombeteiro zonzeira zonzo zote zumbidor

A.3.2 Termos positivos

à-vontade abalizado abalroar abanar abastado abater abençoado aberto abnegado abonado abonatório aborrecer abraçado abrilhantado abrir a alma abrir o coração abrir os olhos absorvente abstémico abstémio acatado aceitável acertado acertar na mosca acertar na mouche acessível achegado acolhedor aconchegado acorrentar acossar acreditado activo acudir acurado adepto adequado adestrado admirável adorado adorando adorável afabilidade afamado afável afectividade afectivo afecto afectuoso afeiçoado afetividade afetivo afeto afetuoso afinado afirmativo afoiteza afoito afortunado afrontar afugentar afundar agarrar a oportunidade com unhas e dentes agarrar a oportunidade agarrar o touro pelos cornos ágil agilidade agir de boa-fé agradado agradar agradável agrado aguentar-se bem aguentar-se airoso ajudado ajuizado alegrar-se alegre alegria alentado alentador alento alicerçado alinhado almo altamente altruísmo altruísta altruístico alumiado amabilidade amachucar amado amadurecer amadurecido amável amedrontar ameno amestrado amigalhão amigável amigo amistoso amorável amorosidade amoroso andar nas sete quintas anestesiador angelical angélico animação animado animador animal político animar animoso aniquilar antidogmático antifascista antinazista anular apagar apaixonante apaixonar apavorar apaziguador apaziguar apelativo apessoado aplicado apolíneo apostado aprazível apreciado apreciador apreciável apresentar alternativas apresentar resultados apresentável aprimorado apropriado aprovado aprumado aprumo apto apurado ardente arnáz arrasar arrebatador arreliar arretado arrojado arrumar a um canto arrumar artístico asfi-

xiar asseado asseio assertivo assiduidade assíduo assinalado assinalável assisado assombrar
atacar atarantar atemorizar atencioso atento aterrorizar atilado atinado ativo atleta atlético
atordoar atormentar atractivo atraente atrair atrapalhar atrativo aturdir áureo auspicioso
autenticidade autêntico autoconfiança autoconfiante autodidacta autodidata autonomia
autonomista autónomo autorizado autosuficiência autosuficiente avançado avasalar
avisado bacano banir batalhador batalhante batalhar bater beleza belo bem disposto
bem vestido bem-amado bem-apesoado bem-aventurado bem-colocado bem-comportado
bem-criado bem-educado bem-encarado bem-falante bem-formado bem-humorado bem-intencionado
bem-parecido bem-pensante bem-posto bem-sucedido bem-vindo bendito benéfico benemerência
benemérito benevolência benevolente benévolo benêvolo benfazejo benfeitor benfeitoria
benignidade benigno benquisto bestialidade boa apresentação boa criação boa disposição
boa educação boa formação boa aparência boa performance boas intenções bom aluno bom
cheiro bom comportamento bom desempenho bom humor bom-samaritano bom bombardear
bondade bondoso boniteza bonito braceiro bravura brilhante brilhar brilho brio briosidade
brioso calar calipígio calma calmante calmo caloroso capacitado capaz captar eleitorado
caricioso caridoso carinho carinhoso carisma carismático caritativo carola castigo castigar
casto castrar categórico cativante cativar cauteloso cauto cavalheiresco cavalheirismo
cavalheiro célebre celestial certo nas contas certo chamar os bois pelos nomes chamativo
charme charmoso cheiroso chique ciente circunspecto cívico civilizado clarificador
clarividência clarividente clemência clemente coerência coerente coeso colocar o dedo na ferida
combater o desemprego comedido compassivo competência competente compincha complacência
complacente complicar a vida a composto compreensão compreensível compreensivo comunicativo
conceituado concertado conciliador conciso concordante concreto confiança confiante confiável
confortador confortante confrontar congruência congruente conhecedor conquistar terreno a conquistar
consagrado consciência consciencioso consciente cômico conseguir um grande feito consentâneo
consequente consertado conservado considerativo consistente conspícuo constante construtivo
consumado contente contrapor contrariar controlado controlar conversa fiada convicção
convitativo convincente convivente cooperador cooperante coragem corajoso cordato cordial
cordialidade correção correcção correcto correto corrigir cortar o pio a cortês cortesão
cortesia cosmopolita cosmopolítico cotado craque credibilidade credível crescido criatividade
criativo crisólogo crisóstomo crispar cristalino criterioso crível crucificar cuidado cuidadoso
cultivado culto cultura democrática cumpridor curado curial curtido dadivoso dar alegrias a dar
cabo de dar conta do recado dar no duro dar o litro dar o primeiro passo dar um baile a dar um
banho a dar um banho de cultura

a dar um show de bola dar uma banhada a dar uma coça a dar uma lição a dar uma lição de moral a dar uma malha a dar uma mãozinha a dar uma surra a dávida debilitar decapitar decência decente decepar decidido decoro decoroso dedicação dedicado defensor deferente deferido deitar água na fervura deleitoso delicadeza delicado delicioso democracia democraticidade democrático denodado derrear derreter derrotar derrubar desacreditar desafectado desafetado desafojado desancar desarmante desarmar descabelar descerimonioso descomplexado descontraído desejável desembaraçado desempenado desenrascado desenhencilhado desenvolvimento desenvoltura desforrar-se desfrutado desinibido deslumbrante deslumbrar desmascarar despachado desperto despretensioso destemido destronar destruir desvelado desvelo detalhado determinação determinado devastar diáfano dialogante dificultar a vida a dificultar dignidade dignificar digno de confiança digno dilacerar diligência diligente diminuir dinâmico dinamismo dinamizador diplomacia diplomático directo direito de contas direito direto disciplina disciplinado discreto descrição distinto ditoso divertido divertimento divinal divino dizer a verdade dizer algumas verdades a dizer umas quantas verdades a dizimar doce dócil docilidade doçura dominador dominante dominar donairoso dotado douto doutrinado durázio eclético eclético ecológico econômico económico economizador edificante educação educado educador educativo eficácia eficaz eficiência eficiente efusivo egrégio elástico electrizante elegância elegante eletrizante eliminar elogio elogioso eloquência eloquente elucidar elucidativo emancipado emblemático emérito empenhado empenho empolgado empreendedor enaltecido enamoramento encaminhado encantador encantar encanto encher as medidas de encolerizar encorajado encorajador encorajamento encostar à parede encostar a um canto encostar às cordas encostar às tábuas encurralar endinheirado endireitar-se energia enérgico enervar enfrentar o touro pelos cornos enfrentar enfurecer engaiolar engraçado enobrecer enraivecer ensombrar entalar entendido enternecedor enternecido enterrar entregar-se de corpo e alma entusiasmante entusiasmar entusiasmo entusiasta entusiástico enxotar enxovalhar enxuto equânime equanimidade equidade equilibrado equilíbrio equitativo equitativo erguer a cabeça erudito esbelto esbofetear esclarecedor esclarecer esclarecido escolarizado escorreito escrupuloso escultural esforçado esfuziante esganar esgotar esmagar esmerado esmerar-se esmero esmoler espancar especial espectacular espectral esperançado esperançoso espetacular espevitado espezinhar espirituoso esplendente esplêndido esplendoroso espontaneidade espontâneo estabilidade estar à altura de estar à altura estar aí para as curvas estar bem estar de saúde estar em grande forma estar em grande estar nas sete quintas estar no bom caminho estarrecer estável estimado estimável estorvar estrangular estrénuo estudioso estupendo etéreo eufónico eufórico exato exceder excelência excelente

excelso excepcional excitante excitar exemplar exemplaridade exímio expansivo expeditivo expedito experiência experiente experimentado experto expressivo expulsar extenuar exterminar extraordinário extremoso extrovertido exultante fabuloso faceto factual falado falante falar com convicção falar verdade famoso fantástico fascinante fascinar favorável fazer boa figura fazer com os olhos fechados fazer com uma perna às costas fazer das tripas coração fazer sombra a fazer um grande feito fazer um negócio da China fazer uma boa jogada fazer uma figura decente fazer uma jogada de mestre fazer xeque-mate fecundo felicidade feliz felizardo fenomenal fértil festivo fiabilidade fiável ficar em primeiro fidedignidade fidedigno fidelidade fiel filantrópico filantropo fintar firme firmeza fixe flagelar flexibilidade flexível fluência fluente fluidez fluido fofo formidável formoso formosura fortalecido forte fortificado fotogenia fotogénico fragilizar franco franqueza fraternal fraternidade fraterno frescalhote frontal frontalidade frugal funcional fundamentado fustigar fuzilar gabaritado galante galhardo ganhar garboso generosidade generoso genial genialidade génio gentil gentileza genuinidade genuíno glamoroso glamour glória glorificado glorioso gostável gostoso governado gozar grácil graciosidade gracioso grandiloquente grandíloquo grandiosidade grandioso grato grimpó guapo habilidade habilidoso habilitado harmonia harmônico harmónico harmonioso hercúleo heroicidade heróico higiénico hílar hiperculto hiperdotado hodierno homogéneo honestidade honesto honorabilidade honorável honra honradez honrado honrar a sua palavra honrar os seus compromissos honrar honroso hospitaleiro hospitalidade humanidade humanitário humano humilhar ideal idolatrado idoneidade idôneo idóneo ileso ilibado iluminado iluminar ilustrado ilustre imaculabilidade imaculado imagético imbatível imolado imortal imortalizado impagável imparcial impávido impecabilidade impecável impertérrito imperturbado imperturbável impoluto imponente impressionar inatacável inaudito incansável incensurável ínclito incólume incomodar inconcusso incontestável incontroverso incorrupção incorruptibilidade incorruptível incorrupto incriticável indefectível indefetível indemnizado inderrotável inderrubável indestrutível indiscutível indispensável indubitável indulgente inelutável inequívoco inesgotável inexcedível infalível infatigável infernizar influente inibir inocente inovação inovador inquestionável insigne inspiração inspirado inspirador inspirar confiança instruído instrutivo insubornável insubstituível insuperável insuspeito integridade íntegro inteligência inteligente intemerato interessante intimidar intimorato invencível inventivo inventor invulnerável ir à luta ir às nuvens irrefutável irrepreensível irreprimível irresistível isento jeitoso jovial jovialidade jubilado jubiloso juízo juntar o útil ao agradável justiça justiceiro justificável justo labutador laudativo laudatório lavado leal lealdade ledó legal legítimo lendário lenitivo lesto letrado levantar a cabeça levar a melhor levar a sua avante lhano

liberdade libertador libertar lídimo limpar o bom nome límpido limpo lindo lisonjeador lisonjeiro lítico livre lógico longânime longânimo louvável lucidez lúcido luminosidade luminoso lustrado lutador lutar luxuoso maduro magistral magnanimidade magnânimo magnificência magnificante magnificidade magnífico mago majestático majestoso maleável malhar maneira maravilha maravilhoso marcar pontos martirizar másculo massacrar matar dois coelhos com uma cajadada maternal maturidade mavioso medrado meigo melífero melódico melodioso memorável memorioso merecedor meritório mesurado meter dentro de um púcaro meter mãos à obra meter num bolso meter num chinelo metucioso metódico mexido milagroso mimoso minucioso misericordioso mítico mociço moderação moderado modernação moderno modéstia modesto moldável monopolizar morigerado mostrar atitude motivação motivado motivador munificente musculado musculoso mutilar não se sair mal não ter papas na língua não virar a cara nascer com o cu virado para a lua natural necessário nítido nobre nomeado normal norteador notabilidade notável objectividade objectivo objetividade objetivo observador observante ofuscar olhar nos olhos onipotente onnipresente onnisciente operante operoso opíparo oportuno oprimir optimismo optimista óptimo ordeiro ordenado organizado original originalidade otimismo otimista ótimo pacato paciente pacificação pacificador pacificidade pacífico paixão paralisar parcimónia parcimonioso parco participativo passar à frente a passável pastel paternal patriarcal patriota patriótico penetrante perceptível perceptivo perçetível perçetivo peregrino peremptório perentório perfeição perfeito perfumado perito perseverança perseverante persistente perspicácia perspicaz perspicuidade perspícuo persuasivo pertinência pertinente perturbar piedoso pitoresco plácido plausível polivalente ponderação ponderado ponderoso pontapear pontual pontualidade popular popularidade porreiro portentoso positividade positivo possante potente poupado pragmático pragmatismo prático prazenteiro precatado precavido precioso precisão preclaro predilecto predileto preeminente preferido premiado prendado preparado presciente presente preservado prestabilidade prestante prestativo prestável prestigiado prestigiante prestígio prestigioso prestimoso presto prevenido providente prezado prezável primoroso probo prodigioso produtividade produtivo proeminente proficiente profícuo profissional profissionalismo progressista prolífero prolífico prometer promissor promitente propício proporcional prosperar prosperidade próspero protector protetor proveitoso providência providencial providente provocar prudência prudente pujante pulcro pundonoroso punir pureza purificado puro qualificado queimar querido querubínico quisto racional radiante radioso rápido raro razoável real realismo realista realizado rebaixar rebater recatado receptível receptividade receptivo recetível recetividade recetivo recomendado recomendável recomposto reconciliador reconfortado reconstituído

recreativo rectidão rectilíneo recto recuperado recuperável redivivo reelegível reeleito refeito reflectido reflectivo refletido refletivo reflexivo reforçado refrescante refulgente refutar regalado regenerado regenerável região regrado regular reinado reinante rejuvenescido relevante remoçado renascer das cinzas renovado repelir reposto reprimir reputado requintado resoluto resolvido respeitabilidade respeitado respeitador respeitável respeito respeitoso responsabilidade responsável ressabido ressarcido ressuscitado restabelecido restaurado retidão retilíneo reto reverenciado revigorado rico ridente ridicularizar rigor rigoroso rijo risonho robustecido robustez robusto sabedoria saber separar as águas sábio sacrificar-se sacudir sadio safar-se sagaz sage sagrado sair de cabeça erguida sair do armário sair-se bem salutar salvar a honra do convento salvar salvo sanado santificado santo são sapiência sapiente sarado satisfação satisfeito saudável sedoso sedutor seduzir segurança seguro selecto seletto semidivino seminal sensacional sensatez sensato sensibilizar sensível sensual sensualidade sentir-se nas nuvens ser um homem às direitas serenidade sereno seriedade sexy silenciar simpatia simpático simpatizante sinceridade sincero siso sistematicidade sistemático soberbo sobredotado sobressair sobriedade sóbrio sociável sofisticação sofisticado solícito solicitude solidário sólido soltar-se sonante sorridente sossegado sovar suave subir nas sondagens sublimado sublime sublimidade sucesso sumptuoso suntuoso superar superconfiante superdotado superelegância superelegante superior supimpa suplantar supremo surpreendente talento talentoso tangível temperado temporizador ter a cabeça no lugar ter a faca e o queijo na mão ter a razão do seu lado ter a vida facilitada ter as costas largas ter boa consciência ter dois palmos de testa ter ideias ter na mão ter os pés assentes no chão ter os pés bem assentes na terra terno ternura ternurento testo tino tolerância tolerante tomar conta do jogo tonificado torturar trabalhadeira trabalhador tramar tranquilidade tranquilizador tranquilizante tranquilo tranquilo translúcido transparência transparente transtornar traquejado tratado tratável triunfador triunfante triunfo trucidar ubíquo ultrapassar unificador uniformizador unívoco útil valentão valente valentia validade válido valioso valoroso vanguardeiro vanguardista vantagem vantajoso varonil veloz vencedor vencer venerado venerando venerável venturoso ver a luz ao fundo do túnel ver com olhos de ver veracidade veraz verdade verdadeiro veridicidade verídico verosímil verosimilhante versado versátil versatilidade vertical vexar viajado viçosidade viçoso vígil vigor vigoroso viril virtuoso visionário vistoso vitalizador vitimar vitória vitorioso vivaço vivaz viver vivo vivificador vivo voluntariosidade xenófilo zelo zeloso zombar

APÊNDICE B - Lista de eventos do oscar em ordem cronológica e testes com tempo

Evento	Data
Melhor roteiro: Spotlight	28/2/16 22:47
Melhor roteiro adaptado: A grande aposta	28/2/16 22:53
Melhor atriz coadjuvante: Alicia Vikander	28/2/16 23:12
Melhor figurino: Mad Max	28/2/16 23:20
Melhor design de produção: Mad Max	28/2/16 23:24
Melhor cabelo e maquiagem: Mad Max	28/2/16 23:27
Melhor fotografia: O regresso	28/2/16 23:38
Melhor edição: Mad Max	28/2/16 23:40
Melhor montagem: Mad Max	28/2/16 23:46
Melhor edição de som: Mad Max	28/2/16 23:49
Melhor mixagem de som: Mad Max	28/2/16 23:52
Melhor efeitos especiais: Ex Machina	28/2/16 23:59
Melhor curta de animação: A História de um Urso (Bear Story)	29/2/16 0:10
Melhor animação: Divertida Mente	29/2/16 0:12
Melhor ator coadjuvante: Mark Rylance	29/2/16 0:32
Melhor documentário curta: A Girl in the River: The Price of Forgiveness	29/2/16 0:41
Melhor documentário: Amy	29/2/16 0:43
Dave Grouhl. Show emória das perdas no cinema em 2015	29/2/16 1:00
Melhor curta: Stutterer	29/2/16 1:07
Melhor filme estrangeiro: O Filho de Saul	29/2/16 1:09
Discurso do Vice presidente sobre abuso sexual nas universidades	29/2/16 1:12
Apresentação Lady gaga	29/2/16 1:15
Melhor trilha sonora: Os Oito Odiados	29/2/16 1:22
Melhor musica original: Sam Smith	29/2/16 1:28
Melhor trilha sonora: Ennio Morricone	29/2/16 1:34
Melhor diretor: Alejandro González Inharrito	29/2/16 1:38
Melhor atriz: Brie Larson	29/2/16 1:48
Melhor ator: Leonardo DiCaprio	29/2/16 1:55
Melhor filme: Spotlight	29/2/16 2:00

	Teste 1	Teste 2	Teste 3	Teste 4
Positivo	15.517	49.263	47.450	69.070
Negativo	17.350	35.079	7210	33.461
Neutro	108.931	57.456	87.138	39.267
Tempo (s)	311,6	397,48	709,129	650,97

Tabela B.1: Comparando testes com tempo