

**UNIVERSIDADE DO VALE DO ITAJAÍ
CENTRO DE CIÊNCIAS TECNOLÓGICAS DA TERRA E DO MAR
CURSO DE CIÊNCIA DA COMPUTAÇÃO**

**ANÁLISE DE SENTIMENTO UTILIZANDO PROCESSAMENTO DE
LINGUAGEM NATURAL**

por

Pablo Fagundes Wachsmann

Itajaí (SC), outubro de 2017

**UNIVERSIDADE DO VALE DO ITAJAÍ
CENTRO DE CIÊNCIAS TECNOLÓGICAS DA TERRA E DO MAR
CURSO DE CIÊNCIA DA COMPUTAÇÃO**

**ANÁLISE DE SENTIMENTO UTILIZANDO PROCESSAMENTO DE
LINGUAGEM NATURAL**

Área de Inteligência Artificial

por

Pablo Fagundes Wachsmann

Relatório apresentado à Banca Examinadora do
Trabalho Técnico-científico de Conclusão de Curso
do curso de Ciência da Computação para análise e
aprovação.

Orientador(a): Rafael de Santiago, MsC

Itajaí (SC), outubro de 2017

AGRADECIMENTOS

Agradeço a minha família, que me deu a estrutura necessária para que eu pudesse enfrentar fases como esta, sempre me incentivando e estando ao meu lado. Aos amigos, Antônio e Yan, pela parceria e todos os momentos divididos, pois apesar de todas as dificuldades, foram anos de muitas conversas e risadas.

A todos que contribuíram de alguma forma para que este trabalho fosse possível, mesmo que com paciência. A todos os professores, que ao longo do curso, nos deram conhecimentos importantes não somente para a formação acadêmica e profissional, mas para a vida. Em especial ao professor Rudimar, que fez correções e sugestões que contribuíram bastante para o trabalho. Aos alunos do laboratório LIA, Rudson, pela ajuda com as classificações das frases. Ao Ivan, pela importante contribuição na pesquisa e no desenvolvimento do trabalho.

Em especial, a Gabriella, que prontamente disponibilizou seu tempo e conhecimento para ajudar com uma parte de grande importância ao trabalho, seu conhecimento técnico foi fundamental para que os resultados do trabalho tomassem forma. E ao professor Santiago, que apoiou a ideia, incentivou, contribuiu e guiou o trabalho, pois sem sua ajuda, a conclusão deste trabalho não teria sido possível.

“A única coisa que temos de respeitar, porque ela nos une, é a língua.” Franz Kafka.

RESUMO

WACHSMANN, Pablo F. Análise de Sentimento Utilizando Processamento de Linguagem Natural. Itajaí, 2017. 163 f. Trabalho Técnico-científico de Conclusão de Curso (Graduação em Ciência da Computação) – Centro de Ciências Tecnológicas da Terra e do Mar, Universidade do Vale do Itajaí, Itajaí, 2017.

Com a massa de dados gerada por usuários na Internet, se faz cada vez mais necessário o uso de ferramentas que sejam capazes de processar estes documentos e extrair informações úteis. Com a capacidade de fornecer métodos capazes de interpretar e processar textos escritos em linguagem humana, as técnicas de Processamento de Linguagem Natural (PLN) despertam o interesse de pesquisadores em diversas áreas. Uma tarefa utilizada em conjunto com o PLN é a análise de sentimento, que pode ser realizada através de diversos métodos e visa extrair o sentimento ou opinião expressada pelo autor de determinado texto. Desta forma, neste trabalho foi desenvolvido uma ferramenta que realiza análise de opiniões coletadas a respeito da mudança das normas de inscrição do ENEM 2017, identificando sentimentos positivos, negativos e neutros. A partir do estudo da tarefa de análise de sentimento, seus métodos e utilização conjunta com o PLN. Para isto, realizou-se: (i) uma análise do funcionamento e uso do PLN; (ii) se verificou a utilização do PLN associado as técnicas de análise de sentimento; (iii) foi selecionado as técnicas a serem utilizadas para o desenvolvimento; (iv) foi especificado o protótipo a ser desenvolvido; (v) foi desenvolvido o modelo especificado; e (vi) o modelo foi testado e comparado com a análise feita por dois alunos do laboratório LIA e um especialista, para que se pudesse obter o nível de qualidade alcançada com o experimento. Com a análise dos resultados, foi possível atingir o objetivo previsto e concluir que o sistema é capaz de classificar as opiniões melhor do que uma pessoa que não possui conhecimentos específicos em linguística.

Palavras-chave: Processamento de Linguagem Natural. Análise de Sentimentos. Análise Textual.

ABSTRACT

With the massive data generated by users on the Internet, it becomes increasingly necessary to use tools that are capable of processing these documents and extract useful information. With the ability to provide methods capable of interpreting and processing texts written in human language, the techniques of Natural Language Processing (NLP) arouse the interest of researchers in several areas. A task used with NLP is the sentiment analysis, which can be performed through several methods and aims to extract the sentiment or opinion expressed by the author of a given text. In this way, in the present work was developed a tool that performs analysis of opinions collected regarding the change of ENEM 2017 registration standards, identifying positive, negative or neutral sentiments. From the study of the task of sentiment analysis, its methods and joint use with NLP. For this, it was carried out: (i) an analysis of the operation and use of the NLP; (ii) it was verified the use of NLP associated with the techniques of sentiment analysis; (iii) the techniques to be used for development were selected; (iv) the prototype to be developed was specified; (v) was developed the specified model; and (vi) the model was tested and compared with the analysis done by two students of the LIA laboratory and a specialist, so that one could obtain the level of quality achieved with the experiment. With the analysis of the results, it was possible to reach the intended objective and to conclude that the system is able to classify the opinions better than a person who does not have specific knowledge in linguistics.

Keywords: Natural Language Processing. Sentiment Analysis. Textual Analysis.

LISTA DE ILUSTRAÇÕES

Figura 1.	Expressão regular para identificar textos em letra maiúscula.	20
Figura 2.	Expressão regular para identificar sequências numéricas.	20
Figura 3.	Relação entre expressão regular, linguagem regular e autômato finito.	21
Figura 4.	AF que representa a conjugação do verbo fazer no indicativo presente.	21
Figura 5.	Resultados das entradas do autômato representado na Figura 4.	22
Figura 6.	POS da frase “As ruas estavam escuras”.	25
Figura 7.	Árvore sintática da frase “o gato caçou o rato”.	29
Figura 8.	Exemplo de árvore de parser.	30
Figura 9.	Exemplo de árvore de parser com anexo semântico.	32
Figura 10.	Técnicas utilizadas para a tarefa de AS.	39
Figura 11.	Representação de um hiperplano da técnica de SVM.	41
Figura 12.	Modelo do sistema de análise de Ko e Seo.	47
Figura 13.	Modelagem do sistema de Kim e Hovy.	50
Figura 14.	Modelagem do sistema proposto.	55
Figura 15.	Fluxo de operação do sistema proposto.	57
Figura 16.	Logo da biblioteca CoGrOO, utilizada no módulo PLN.	59
Figura 17.	Exemplo de identificação de lemas e classes.	59
Figura 18.	Tela principal do sistema desenvolvido.	60
Figura 19.	Exemplo de frase com opiniões identificadas.	61
Quadro 1 -	Exemplos de palavras etiquetadas e suas características	27
Quadro 2 -	Exemplo de etiquetagem de frase para árvore de parser.	30
Quadro 3 -	Exemplo de gramática semântica.	33
Quadro 4 -	Exemplo de caso.	34

LISTA DE TABELAS

Tabela 1 -	Fatores de saliência	35
Tabela 2 -	Complemento de fatores de saliência.....	36
Tabela 3 -	Quantidade de adjetivos e verbos utilizados	51
Tabela 4 -	Adjetivos inseridos manualmente no dicionário	61
Tabela 5 -	Resultados obtidos na análise das opiniões pela a especialista.....	63
Tabela 6 -	Resultados obtidos na análise das opiniões identificadas pelo Aluno 1	64
Tabela 7 -	Resultados obtidos na análise das opiniões identificadas pelo Aluno 2	65
Tabela 8 -	Resultados obtidos na análise das opiniões classificadas pelo Aluno 1 comparando com as identificadas pela Especialista.	66
Tabela 9 -	Resultados obtidos na análise das opiniões classificadas pelo Aluno 2 comparando com as identificadas pela Especialista.	67

LISTA DE ABREVIATURAS E SIGLAS

ABT	Aprendizado Baseado em Transformação
AF	Autômato Finito
AS	Análise de Sentimento
CBA	Classificação Baseada em Associações
CNB	Classificador Naive Bayes
ER	Expressão Regular
ML	Machine Learning
MO	Mineração de Opinião
PLN	Processamento de Linguagem Natural
POS	Part-of-Speech
RAC	Regra de Associação de Classe
RBF	Radial Basis Function
RNA	Rede Neural Artificial
SVM	Support Vector Machine

SUMÁRIO

1 INTRODUÇÃO.....	11
1.1 Problematização.....	13
1.2 Formulação do Problema.....	13
1.2.1 Solução Proposta.....	14
1.3 Objetivos.....	15
1.3.1 Objetivo Geral.....	15
1.3.2 Objetivos Específicos.....	15
1.4 Metodologia	15
1.5 Estrutura do Trabalho.....	16
2 FUNDAMENTAÇÃO TEÓRICA	18
2.1 PROCESSAMENTO DE LINGUAGEM NATURAL.....	18
2.1.1 Análise Morfológica.....	18
2.1.2 Análise Sintática.....	22
2.1.2.1 Part-of-speech Tagging.....	25
2.1.2.2 POS Baseado em Regras.....	26
2.1.2.3 POS estocástico (probabilístico).....	27
2.1.2.4 POS Baseado em Transformação.....	28
2.1.2.5 N-grams.....	28
2.1.2.6 Árvores de Parser.....	29
2.1.3 Análise Semântica	31
2.1.3.1 Gramática Semântica	32
2.1.3.2 Gramáticas Baseadas em Caso.....	33
2.1.4 Análise Pragmática	34
2.2 ANÁLISE DE SENTIMENTO	37
2.2.1 Machine Learning	39
2.2.1.1 Supervisionado	40
2.2.1.2 Não-Supervisionado	46
2.2.2 Léxico	49
2.2.2.1 Uso do PLN em Conjunto com métodos Léxicos.....	53
3 DESENVOLVIMENTO	54
3.1 PROJETO	54
3.2 DESENVOLVIMENTO	58
3.3 Resultados	62
4 CONCLUSÕES.....	69

1 INTRODUÇÃO

Segundo Sun, Luo e Chen (2016), com o grande crescimento de textos gerados por usuários na Internet, a extração automática de informação útil de tantos documentos, desperta interesse de pesquisadores em diversas áreas, em particular na área de Processamento de Linguagem Natural (PLN). Ao contrário de linguagens artificiais, como linguagens de programação e notações matemáticas, linguagens naturais evoluem com o passar das gerações e são difíceis de se definir com regras explícitas. O PLN, oferece ferramentas para a manipulação de linguagens naturais, podendo ser simples como contagens de frequência de palavras para comparar diferentes estilos de escrita, como o completo entendimento de enunciados humanos (BIRD; KLEIN; LOPER, 2009).

De acordo com Covington (1994), o PLN define-se como o uso de computadores para entender linguagens humanas. Não no sentido de um computador poder expressar sentimentos ou pensamentos, mas de maneira que um computador possa reconhecer e usar informação expressada em linguagem humana. Ainda segundo Covington (1994), a estrutura de qualquer linguagem se divide em cinco níveis: fonológica, que é expressada através dos sons; morfológica, interpretada através das formações das palavras; sintática, levando em consideração a estrutura das sentenças; semântica, com foco nos significados das palavras em si; e pragmática, representando o uso da linguagem dentro de um determinado contexto.

O PLN, visa o desenvolvimento de ferramentas para se executar as tarefas necessárias no pré-processamento de um texto a ser analisado. Segundo Sun, Luo e Chen (2016), etapas de pré-processamento são fundamentais para qualquer análise mais detalhada sobre algum tema específico, para se estruturar o texto a ser analisado e extrair suas características. Tarefas como *tokenization* (separação de uma sequência de caracteres separados por *tokens*), desambiguação e *part-of-speech tagging* (designar categorias para cada palavra identificada), são de fundamental importância para que posteriormente possa ser implementado uma análise de sentimento do texto estudado.

Segundo Devika, Sunitha e Ganesh (2016), a Análise de Sentimento é um processo intelectual de extração de sentimentos e emoções de usuários, sendo um dos campos que procedem o PLN. É o processo de detecção de polaridade em textos, bem como determinar se a opinião de um dado texto é positiva, negativa ou neutra.

A Análise de Sentimento (AS) ou Mineração de Opinião (MO), é o estudo computacional de opiniões, atitudes e emoções de acordo com uma entidade, sendo que esta entidade pode ser representada por indivíduos, eventos ou tópicos. As expressões Análise de Sentimento e Mineração de Opinião possuem o mesmo significado, entretanto alguns pesquisadores apontam uma pequena diferença entre elas. A MO extrai e analisa opiniões pessoais sobre uma entidade, enquanto a AS identifica um sentimento expressado em um texto e então o analisa. Portanto, o objetivo da AS é encontrar opiniões, identificar o sentimento que essas opiniões expressam e então classificar sua polaridade (MEDHAT; HASSAN; KORASHY, 2014).

Ainda segundo Medhat, Hassan e Korashy (2014), em AS existem três níveis principais de classificação. O primeiro é o nível de documento, que visa classificar a opinião de um documento e expressar um sentimento positivo ou negativo a seu respeito. Considera todo o documento como base de informação. Em seguida há o nível de sentença, tendo como objetivo classificar o sentimento expressado em cada sentença. Identifica se a sentença é objetiva ou subjetiva e no caso de uma sentença ser subjetiva, determina se a mesma expressa uma opinião positiva ou negativa. Finalmente, o nível de aspecto tem como objetivo classificar o sentimento que diz respeito aos aspectos específicos da entidade. Identifica os aspectos de cada entidade e possui detentores de opinião, que podem estabelecer uma opinião diferente para cada aspecto de uma mesma entidade.

Existem diversos métodos para se realizar AS, porém diversos pesquisadores continuam a procurar melhores soluções, devido a sua importância. A abordagem por Machine Learning funciona treinando um algoritmo com um conjunto de dados de treinamento antes de aplicar no conjunto de dados em si. Estas técnicas, primeiramente treinam os algoritmos com dados específicos de entrada, que possuem saídas conhecidas, para que depois possam ser aplicados em dados desconhecidos (DEVIKA; SUNITHA; GANESH, 2016).

De acordo com Sun, Luo e Chen (2016), abordagens Machine Learning representam um papel significativo na análise de sentimento. Geralmente análises a nível de documento e sentença podem ser formuladas como problemas de classificação, com o objetivo de determinar um sentimento positivo ou negativo, desta forma, pode-se treinar classificadores para determinar a polaridade dos problemas. Métodos de Machine Learning como o

classificador Naive Bayes ou Support Vector Machine são alguns dos modelos mais utilizados.

Segundo Devika, Sunitha e Ganesh (2016), podem-se destacar ainda as abordagens baseadas em regras, que são utilizadas definindo-se várias regras para se extrair as opiniões, criando-se *tokens* para cada sentença em cada documento e então testando-se cada *token*; e as abordagens léxicas, que funcionam assumindo que a polaridade coletiva de um documento ou sentença é a soma das polaridades das frases ou palavras individuais.

Levando em consideração esta contextualização, o presente trabalho propôs o desenvolvimento de um software para detecção de sentimentos em textos. Para isto realizou-se: (i) uma análise do funcionamento e uso do PLN; (ii) verificou-se a utilização do PLN associado a técnicas de Análise de Sentimento; (iii) foi selecionado as técnicas a serem utilizadas para o desenvolvimento; (iv) especificou-se o protótipo a ser desenvolvido; (v) realizou-se a implementação de acordo com as especificações; e (vi) se analisou a qualidade obtida pelo produto da pesquisa (a solução PLN).

1.1 PROBLEMATIZAÇÃO

1.2 FORMULAÇÃO DO PROBLEMA

Tentativas de desenvolver modelos matemáticos para se lidar com linguagem natural tem um longo histórico (TARSKI, 1944 apud RUS, 2002). A necessidade da modelagem da linguagem humana se tornou ainda mais importante com a globalização (ZADROZNY et al 2000 apud RUS, 2002).

Com o grande crescimento de textos gerados por usuários na Internet, a extração automática de informação útil a partir desta quantidade substancial de documentos, recebe interesse de pesquisadores de diversas áreas, e em particular da comunidade de PLN. A análise de sentimento foi proposta a princípio no início do século, se tornando uma área de pesquisa ativa gradualmente (SUN; LUO; CHEN, 2016).

Encontrar opiniões e sentimentos em sites e monitorá-los através da Web pode ser uma tarefa difícil de ser realizada devido ao grande volume de informações. Sentimentos e opiniões podem estar escondidos em longos posts de fóruns ou blogs, e é difícil para uma

pessoa encontrar sites relevantes, extrair sentenças relacionadas a opiniões, lê-las e organizá-las de forma útil (LIU, 2010).

Ainda segundo Liu (2010), em um sistema de análise de sentimento é importante saber o que se quer extrair de um texto. Textos gerados por usuários, podem conter não apenas uma, mas várias opiniões. Podem expressar um sentimento a respeito do alvo da análise de uma forma geral ou de características específicas do mesmo, que talvez não sejam relevantes para a análise que se quer realizar. Em um texto, deve se identificar o detentor de opinião, sendo a pessoa ou organização que está expressando a opinião; a opinião sobre uma característica c , sendo uma visão positiva, negativa ou uma avaliação de c de um determinado detentor de opinião. Desta forma, pode-se modelar o objeto, o texto que expressa sentimentos e o objetivo de análise do texto, gerando um modelo de análise de sentimento baseado em características.

1.2.1 Solução Proposta

A seguinte proposta prevê o estudo das ferramentas para se realizar PLN, utilizadas nas etapas de pré-processamento de textos a fim de prepará-los para serem analisados. Pesquisar métodos existentes para se realizar análise de sentimento. Desenvolver um software que utilize essas ferramentas e técnicas em conjunto para efetuar a AS, identificando os sentimentos positivos, negativos e neutros dentro dos textos. E validar a qualidade da implementação através da assertividade da análise de sentimento em opiniões coletadas na Internet através de um formulário Google.

A pesquisa será realizada com base nas seguintes perguntas:

1. Quais são as técnicas mais utilizadas para o processamento de linguagem natural?
2. Quais os métodos mais utilizados para se fazer Análise de Sentimento em textos?
3. Que base de dados pode ser utilizada para que se extraia informações relevantes?
4. Qual linguagem de programação oferece melhor suporte para implementação?
5. Quais sentimentos podem ser extraídos na análise de um documento?
6. Que sentimentos são mais relevantes para a análise?
7. Qual a qualidade das análises de trabalhos similares?

1.3 OBJETIVOS

1.3.1 Objetivo Geral

Desenvolver um software que realize a análise de sentimento de um texto inserido na língua portuguesa.

1.3.2 Objetivos Específicos

1. Realizar fundamentação teórica a respeito do uso e funcionamento do processamento de linguagem natural e seu uso para a realização da tarefa de análise de sentimento, analisando trabalhos similares a fim de esclarecer as possíveis estratégias de implementação;
2. Definir a estratégia de implementação a ser utilizada para o desenvolvimento do projeto.
3. Especificar o software a ser desenvolvido.
4. Implementar as especificações;
5. Analisar a implementação realizada, comparando com os resultados obtidos através da análise de dois alunos do Laboratório de Inteligência Aplicada da Univali e de um especialista;
6. Ao final do experimento, obter um nível de qualidade satisfatório de acordo com a literatura e trabalhos similares.

1.4 METODOLOGIA

O presente trabalho buscou por meio de pesquisa bibliográfica, obter conhecimento a respeito das ferramentas de PLN, utilizadas para pré-processar textos em computadores, a fim de prepará-los para outras tarefas de análise posteriores. Nesta etapa, fez-se claro a estrutura e o uso da linguagem humana através das etapas de análise estudadas, a fim de processá-la do ponto de vista da máquina.

Posteriormente foi realizada uma pesquisa para identificar a relação do PLN com a análise de sentimento. Explorando, também bibliograficamente, a tarefa de análise de sentimento, seu funcionamento, uso, técnicas mais utilizadas e aplicação. Nesta etapa, são apresentadas as descrições de trabalhos similares.

Com o conhecimento obtido até então, foi especificado o projeto a ser desenvolvido. A linguagem de programação a ser utilizada para o desenvolvimento, a técnica a ser utilizada para o PLN, o método a ser implementado para realizar a análise de sentimento e as ferramentas e bibliotecas que serão utilizadas para apoio.

Por fim, com o conhecimento obtido através da pesquisa bibliográfica efetuada e a especificação do projeto, foi possível desenvolver a solução proposta, comparando os resultados obtidos através da análise das classificações realizadas com a implementação, com as classificações efetuadas por dois alunos do laboratório LIA, bem como um especialista pós-graduado em metodologia da língua portuguesa. Desta forma sendo viável a verificação da qualidade do experimento.

1.5 ESTRUTURA DO TRABALHO

O presente trabalho está dividido em quatro capítulos: Introdução, Fundamentação Teórica, Desenvolvimento e Conclusões. O Capítulo 1, Introdução, apresentou uma visão geral do trabalho, formulando o problema, propondo a solução e apresentando os objetivos gerais e específicos. No Capítulo 2, Fundamentação Teórica, é apresentado a pesquisa bibliográfica sobre:

1. O PLN, onde é descrito todas as etapas de análise que devem ser efetuadas em um texto para que seja possível estruturá-lo, preparando-o para análises posteriores;
2. A AS, sendo apresentado uma visão geral das técnicas utilizadas atualmente e descritos trabalhos similares realizados na área;
3. A relação entre o pré-processamento do texto e as tarefas posteriores de AS.

No Capítulo 3, apresenta-se o desenvolvimento do projeto implementado, incluindo sua especificação, modelagem e fluxo de operações, bem como a proposta de coleta e análise de dados, experimento realizado, análise de apoio realizada por alunos e especialista, e a qualidade obtida ao final do experimento. Por fim, no Capítulo 4, apresenta-se as conclusões,

onde é realizado um apanhado geral do trabalho, avaliando a metodologia utilizada, dificuldades encontradas, as técnicas e ferramentas utilizadas e os trabalhos futuros sugeridos.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 PROCESSAMENTO DE LINGUAGEM NATURAL

O Processamento de Linguagem Natural (PLN), consiste no desenvolvimento de modelos computacionais com objetivo de realizar tarefas de reconhecimento, processamento e compreensão de informações expressas em uma língua natural (PEREIRA, 2007). De acordo com Jurafsky e Martin (2000), para uma máquina compreender e processar uma informação expressa em linguagem natural, é necessário considerar cinco aspectos: fonético, morfológico, sintático, semântico e pragmático. Covington (1994), divide a análise estrutural da linguagem nos mesmos cinco níveis citados acima, porém no presente trabalho, não será abordado a análise do ponto de vista fonético, uma vez que o objetivo do trabalho é o de processar textos já transcritos.

A análise morfológica, consiste no estudo da construção das palavras a partir da unidade de seu menor significado, seus radicais (parte original e irreduzível onde se encontra o significado das palavras) e afixos (partes secundárias que somadas aos radicais formam as palavras) (JURAFSKY; MARTIN, 2000). Enquanto a análise morfológica é realizada a partir da divisão das palavras em suas partes constituintes, a análise sintática tem como foco de estudo as sentenças e suas regras gramaticais, sendo o processo de divisão das palavras que compõem as sentenças (MOTE, 2002). Já a análise semântica, além da ordem estrutural das palavras de acordo com suas regras gramaticais, leva em consideração o significado individual de cada palavra dentro de sua sentença, combinando-as a fim de formar significados mais abrangentes (JURAFSKY; MARTIN, 2000), em outras palavras, a análise semântica é a busca por significado. Segundo Müller (2003), a análise pragmática, último nível de análise linguístico, se atém ao contexto e significado mais amplo das sentenças, a coerência textual e análise de discurso. Buscando nas demais sentenças o entendimento do contexto que falta a frase em análise.

2.1.1 Análise Morfológica

Palavras são a estrutura fundamental da linguagem. Qualquer linguagem natural, é composta por palavras. Toda a área de PLN, desde o reconhecimento da fala até a tradução automática de páginas da web, exige um grande conhecimento sobre palavras (JURAFSKY; MARTIN, 2000). Morfologia é a área da linguística que estuda a estrutura interna destas

palavras, podendo quebrá-las em partes menores que contenham algum significado, preocupando-se com o modo em que estas palavras são compreendidas (ARONOFF, 1976). Se utilizarmos a palavra “cachorrinhos” como exemplo, conseguimos quebrá-la em partes menores, identificando seus elementos, sendo eles:

cachorr – o elemento raiz da palavra, aquele que dá o seu significado.

inh – sufixo indicando que a palavra é um diminutivo

o – sufixo indicando que a palavra está em sua forma masculina

s – sufixo que indica que a palavra se encontra no plural.

Cada unidade mínima destas palavras que possuem significados são chamadas de morfema. Os morfemas se dividem em raiz (ou radical, elementos base e significativos das palavras) e afixos (prefixos, sufixos, desinência ou qualquer elemento modificador da parte significativa raiz).

Uma das linguagens que busca padronização na especificação de *strings* de busca em textos, é a expressão regular (ER). A ER é uma importante ferramenta para a Ciência da Computação e a Linguística. Ela é uma fórmula escrita em uma linguagem especial usada para especificar classes simples de *strings*, onde uma sequência de caracteres é representada por uma sequência de símbolos. Para a maior parte das técnicas de busca que são baseadas em textos, uma sequência de caracteres é qualquer sequência de caracteres alfanuméricos (letras, números, espaços, tabulações e pontuações) (JURAFSKY; MARTIN, 2000).

Ainda segundo Jurafsky e Martin (2000), conceitualmente falando, uma ER é uma notação algébrica para a caracterização de um conjunto de *strings*. Elas podem ser utilizadas para especificar *strings* de busca ou para definir uma linguagem formal. Buscas por ER exigem um padrão desejado a se buscar e textos aonde se deseja realizar a busca, desta forma a função irá passar pelo texto inserido, retornando toda a série de caracteres que contenham o padrão inserido. Nas Figuras 1 e 2, pode-se observar exemplos simples de ER, onde na Figura 1 é descrito uma ER para identificar uma sequência de caracteres maiúsculos e na Figura 2 identificar sequências de caracteres numéricos.

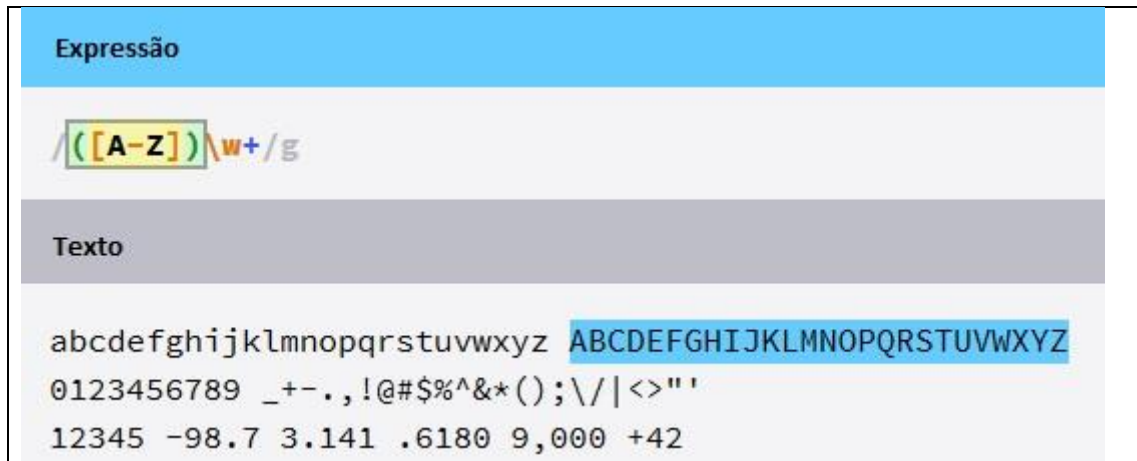


Figura 1. Expressão regular para identificar textos em letra maiúscula.

Fonte: <http://regexr.com/>.

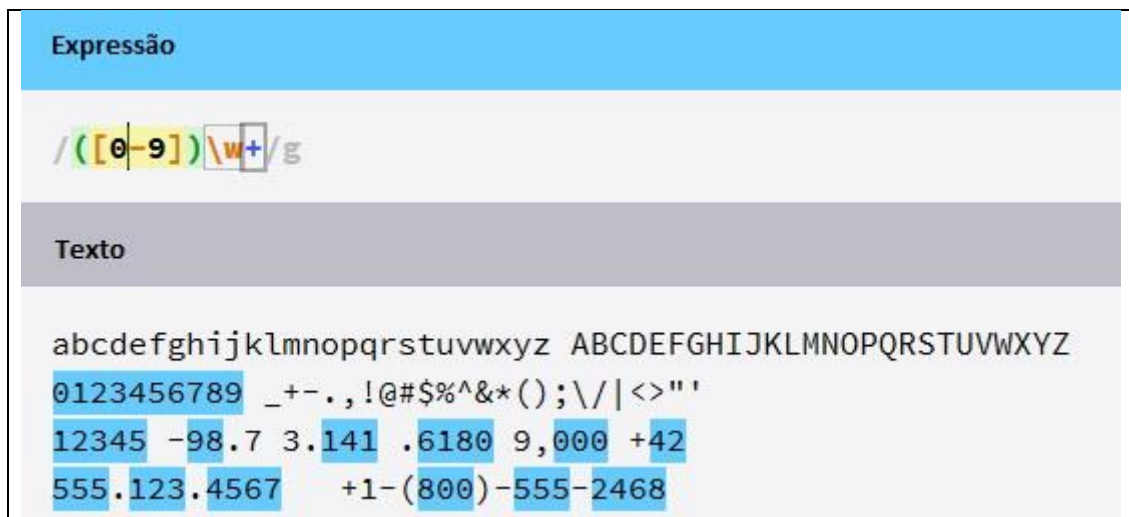


Figura 2. Expressão regular para identificar sequências numéricas.

Fonte: <http://regexr.com/>.

Além de ser uma maneira de caracterizar um tipo de linguagem formal chamada de linguagem regular, como pode ser observado na Figura 3, as expressões regulares são formas de descrever autômatos finitos. Qualquer expressão regular pode ser implementada como um autômato finito e vice-versa (JURAFSKY; MARTIN, 2000). Autômatos finitos (AF), são importantes ferramentas utilizadas em PLN e apesar de serem usados em várias etapas, a nível de análise morfológica são indispensáveis (MOTE, 2002).

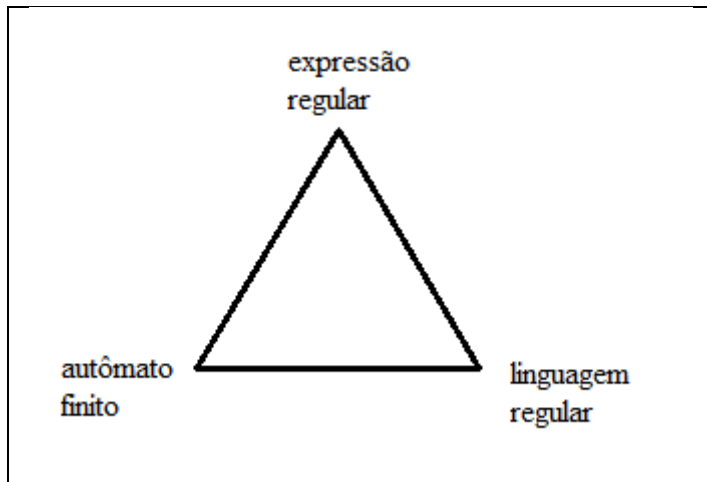


Figura 3. Relação entre expressão regular, linguagem regular e autômato finito.

Fonte: adaptado de Jurafsky e Martin (2000).

Ainda segundo Mote (2002), máquinas de estado como os AF's não são máquinas físicas, não possuem partes mecânicas ou elétricas, mas são descritas como uma ideia abstrata para reconhecer linguagens. A entrada de um AF consiste em um dado de algum alfabeto, podendo este ser um caractere ASCII, palavras ou frases inteiras.

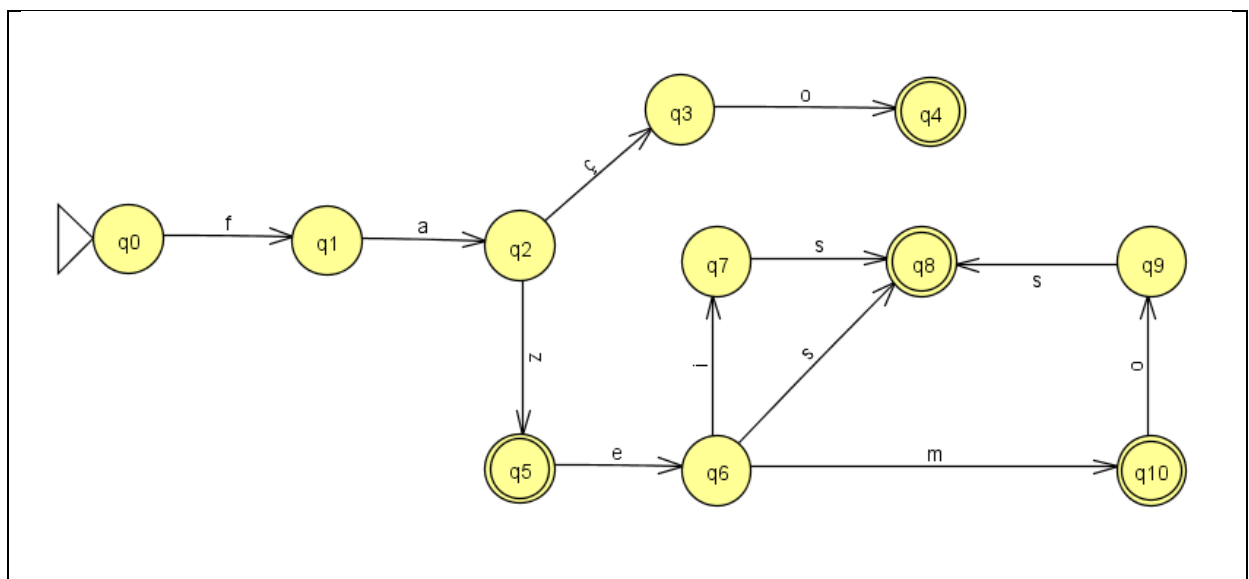


Figura 4. AF que representa a conjugação do verbo fazer no indicativo presente.

Input	Result
faço	Accept
fazes	Accept
fazemos	Accept
fazeis	Accept
fazem	Accept
fazias	Reject
fazom	Reject
fazeu	Reject
fizeu	Reject
fiz	Reject
foz	Reject
fazamos	Reject
aus	Reject
eus	Reject
eis	Reject
foi	Reject
faz	Accept

Figura 5. Resultados das entradas do autômato representado na Figura 4.

O autômato representado na Figura 4, possui onze estados e reconhece a conjugação do verbo fazer na forma do indicativo presente, identificando o conjunto de caracteres um a um. É representado como um grafo direcionado, onde os vértices são cada estado possível do autômato e suas arestas são as transições de cada estado. Na Figura 5, pode se verificar as entradas com seus respectivos resultados após a execução do AF, reconhecendo apenas as palavras que fazem parte da conjugação do verbo mencionado acima.

2.1.2 Análise Sintática

Segundo Jurafsky e Martin (2000), a análise sintática é o estudo das relações formais entre as palavras. Se as palavras são o fundamento do processamento da linguagem, a sintaxe é o esqueleto. Basicamente, a sintaxe tem seu foco de estudo nas sentenças, é o processo de quebra das frases em suas partes constituintes, com o objetivo de analisar principalmente as regras gramaticais que sustentam as sentenças (MOTE, 2002).

As palavras são tradicionalmente agrupadas em classes de equivalência, chamadas de *parts-of-speech* (POS), significando “partes da fala”, iremos utilizar a denominação original em inglês, pois existe uma tarefa de PLN chamada de *part-of-speech tagging* da qual trataremos mais adiante. O POS de uma palavra, não somente de classes principais como verbos e substantivos, mas de outras classes também, nos dá grande informação sobre a palavra em si e suas vizinhas dentro de uma sentença (JURAFSKY; MARTIN, 2000).

Segundo Cegalla (2010), existem dez classes de palavras na língua portuguesa, sendo elas: substantivo, artigo, adjetivo, numeral, pronome, verbo, advérbio, preposição, conjunção e interjeição. Dentre elas, as seis primeiras são classes variáveis, ou seja, podem ser flexionadas e sofrerem alterações em suas formas. Por exemplo, os substantivos podem ser flexionados em número, gênero e grau. As quatro últimas são invariáveis. Os substantivos e verbos se destacam como as duas principais classes de palavras, pois constituem a base das sentenças. Palavras que possuem a mesma classe, tem características em comum. Cada uma das classes possui uma função sintática dentro das frases. Abaixo, é descrito o conceito de cada uma das classes.

Substantivos são palavras que designam os seres e exercem funções sintáticas como sujeito, objeto direto e objeto indireto. Dividem-se em comuns, que designam seres da mesma espécie (menino, galo, palmeira); nomes próprios que se aplicam a um ser em particular (Deus, Brasil); concretos, designando seres de existência real (mulher, homem, leão, pedra); abstratos, que designam qualidades, sentimentos, ações ou estados dos seres (beleza, coragem, amor); simples, formados por apenas um radical (pão, chuva); compostos, sendo formados por mais de um radical (guarda-chuva); primitivos, sendo os que não derivam de outras palavras da língua portuguesa (pedra, ferro); derivados, derivam de outra palavra (pedreiro, ferreiro); e coletivos, que expressam um conjunto de seres da mesma espécie (rebanho, constelação, vara) (CEGALLA, 2010).

Artigos são palavras que antepostas aos substantivos dão aos seres um sentido determinado ou indeterminado. Exemplos de artigos definidos são “o”, “a”, “os” e “as”. Exemplos de artigos indefinidos são “um”, “uma”, “uns” e “umas” (CEGALLA, 2010).

Adjetivos são as palavras que expressam as qualidades ou características dos seres. Exercem as funções sintáticas de predicativo e adjunto adnominal. Dividem-se de acordo com sua formação e podem ser: primitivo, quando não deriva de outra palavra (bom, forte, feliz); derivado, derivando de substantivos ou verbos (famoso, amado); simples, formado por um só elemento (brasileiro); e composto, formado por mais de um elemento (luso-brasileiro) (CEGALLA, 2010).

Numerais são palavras que exprimem número, ordem numérica, múltiplo ou fração, podendo ser cardinal (dois), ordinal (segundo), multiplicativo (dobro) ou fracionário (meio) (CEGALLA, 2010).

Pronomes são as palavras que substituem os substantivos ou os determinam, indicando a pessoa da sentença. Dividem-se em: (i) pessoais, que são as palavras designadas a substituir os substantivos e representar as pessoas do discurso (eu, tu, ele, ela); (ii) possessivos, referem-se as pessoas no discurso, dando-lhes a posse de algo (meu, tua, sua, nosso); (iii) demonstrativos, são utilizados para indicar lugar, posição ou a identidade dos seres (este, esse, aquele); (iv) relativos, palavras que representam substantivos já referenciados, com os quais estão relacionados (que, quem, onde, quanto); (v) indefinidos, referem-se à terceira pessoa do discurso, designando de modo vago, indeterminado (algo, tudo, fulano); e (vi) interrogativos, os pronomes interrogativos são utilizados em sentenças interrogativas e como os indefinidos referem-se de modo impreciso à terceira pessoa do discurso (que, quem, quantos, quê) (CEGALLA, 2010).

Verbos são palavras que exprimem ações, estados, fatos ou fenômenos, são indispensáveis na organização das sentenças. Dentre as classes, o verbo é a que mais possui flexões e seu objetivo é expressar diferentes formas para indicar a pessoa do discurso, seu número, tempo, modo e voz. Ao seu conjunto ordenado de flexões ou formas dá-se o nome de conjugação, e como exemplos podemos citar na forma irregular “dar”, “copiar”, “crer”, “fazer”, “escrever”, etc. (CEGALLA, 2010).

Advérbios são palavras que modificam o sentido do verbo, adjetivo e do próprio advérbio. Nas sentenças exercem a função sintática de adjunto adverbial e em sua maioria modificam o verbo, acrescentando uma circunstância. Apenas os advérbios de intensidade é que podem modificar adjetivos e os próprios advérbios. Podem ser: (i) afirmação (sim, realmente, certamente); (ii) dúvida (talvez, acaso, porventura, decerto); (iii) intensidade (muito, pouco, bastante, mais); (iv) lugar (abaixo, acima, lá, aqui); (v) modo (bem, mal, assim, devagar, como); (vi) negação (não, tampouco); e (vii) tempo (agora, hoje, amanhã, depois) (CEGALLA, 2010).

Preposições são palavras invariáveis que ligam um termo dependente a um termo principal, estabelecendo uma relação entre ambos. Podem ser: essenciais (a, ante, após, até, com, contra); e acidentais (conforme, segundo, durante, como, mediante) (CEGALLA, 2010).

Conjunções são palavras invariáveis que ligam sentenças ou palavras da mesma sentença. As conjunções que ligam as sentenças sem fazer com que uma dependa da outra, e que a segunda sentença não complete o sentido da primeira, são chamadas de conjunções

coordenativas, já as que ligam duas orações que se completam, fazendo com que a segunda dependa da primeira, são conjunções subordinativas (CEGALLA, 2010).

Interjeições são palavras ou locuções que exprimem um estado emotivo (“caramba!”, “meu deus!”, “oh!”) (CEGALLA, 2010).

2.1.2.1 Part-of-speech Tagging

Part-of-speech (POS) Tagging é o processo de “etiquetagem”, atribuindo classes ou outra marcação léxica de classe para cada palavra dentro do texto, estas marcações também são aplicadas para pontuações e em muitos casos podem ser ambíguas. O POS representa uma importante tarefa no processo de análise da linguagem natural e na recuperação da informação. A entrada para um algoritmo de POS é uma *string* de palavras e um conjunto de “etiquetas” especificadas para descrever as classes (JURAFSKY; MARTIN, 2000). Desta forma, se for estabelecido que os verbos serão representados pela etiqueta VB, artigos por AR, adjetivos por AD e pronomes como PR, se teria uma representação semelhante a demonstrada na Figura 6:

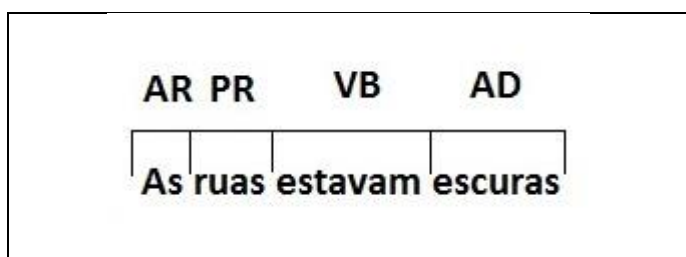


Figura 6. POS da frase “As ruas estavam escuras”.

Fonte: adaptado de Jurafsky e Martin (2000).

Jurafsky e Martin (2000) descrevem que, esta tarefa é importante, pois existem palavras que podem ser ambíguas. O objetivo do POS é resolver ambiguidades etiquetando as palavras de forma correta dentro do contexto da sentença. Segundo Maletti (2017), uma vez que utilizamos o PLN para preparar os textos para a AS ou MO, e o objetivo final é o de encontrar polaridades, ao se realizar o POS, deve-se ater principalmente as classes que por si só possuem conteúdo, como substantivos, adjetivos, verbos e negações. Deixando de lado classes que expressam funções, como artigos, pronomes, verbos auxiliares, pois palavras que se enquadram nestas classes normalmente servem uma função gramatical e não contém significado léxico. Maletti (2017) alerta para a ambiguidade. A palavra da língua portuguesa “bem”, por exemplo, pode fazer parte de pelo menos cinco classes (substantivo, advérbio,

pronome, adjetivo e interjeição), observando que a tarefa de POS é muito importante para se tratar estas ambiguidades.

Jurafsky e Martin (2000), descrevem três tipos principais de algoritmos para se realizar POS: (i) os etiquetadores baseados em regra, que utilizam uma grande base de dados de regras de desambiguações escritas manualmente especificando, por exemplo, que uma determinada palavra é um substantivo ao invés de um verbo; (ii) os etiquetadores estocásticos, que resolvem a ambiguidade usando um conjunto de treinamento para estabelecer a probabilidade de uma determinada palavra possuir determinada etiqueta de classe dentro de seu contexto; e (iii) etiquetadores baseados em transformação que utilizam elementos combinados dos dois métodos citados anteriormente.

2.1.2.2 POS Baseado em Regras.

Os primeiros algoritmos para se realizar POS automaticamente eram baseados em dois estágios. O primeiro estágio, utilizava um dicionário para atribuir potenciais classes de POS a cada palavra, e o segundo usava uma grande lista de desambiguações escritas manualmente, para restringir as potenciais classes em apenas uma só.

O ENGTWOL (Voutilainen, 1995), é um etiquetador baseado nos mesmos dois estágios, porém ambos são mais sofisticados que os algoritmos propostos originalmente. A análise no primeiro estágio é feita com base em uma morfologia de dois níveis e possui cerca de cinquenta e seis mil entradas para palavras em inglês. Ele conta as palavras com múltiplas classes como entradas separadas, ignorando formas flexionadas e derivadas. Cada entrada é anotada como um conjunto de características morfológicas e sintáticas. No Quadro 1 pode-se observar um exemplo de um conjunto de palavras com suas características (JURAFSKY; MARTIN, 2000).

Quadro 1 - Exemplos de palavras etiquetadas e suas características

Palavra	POS	Característica
Menor	ADJ	Adjetivo comparativo
Inteiro	ADJ	Adjetivo atributivo absoluto
Rápido	ADV	Advérbio superlativo
Aquele	DEM	Demonstrativo singular
Todo	DEM	Demonstrativo quantificador
Cachorros	S	Substantivo genitivo singular
Móvel	S	Substantivo nominativo singular
Um-terço	NUM	Numeral singular
Ela	PR	Pronome feminino singular
Ocorreu	V	Verbo no passado

Fonte: adaptado de Jurafsky e Martin (2000).

Após este primeiro processamento, é aplicado um conjunto de cerca de mil restrições no texto de entrada. Para que os POS incorretos sejam descartados, estas restrições são aplicadas na forma negativa, eliminando as etiquetas que não estão de acordo com o contexto. Está é uma descrição simplista do sistema que simula a tarefa de POS, ele ainda conta com regras probabilísticas e faz uso de outras informações sintáticas que não foram abordadas (JURAFSKY; MARTIN, 2000).

2.1.2.3 POS estocástico (probabilístico).

Segundo Jurafsky e Martin (2000), existem vários algoritmos para fazer a tarefa de POS através de etiquetagem probabilística. O POS probabilístico atribui a classe que mais provavelmente se adequa a cada uma das palavras. O Modelo de Markov é o principal método utilizado para esta tarefa. Para uma dada sentença ou sequência de palavras, os etiquetadores de Markov escolhem uma sequência de classe maximizada pela Equação 1:

$$e_i = \operatorname{argmax}_j P(e_j | e_{i-1}, w_i) \quad (1)$$

Desta forma, conforme a Equação 1 o algoritmo de Markov irá escolher a etiqueta e_i para a palavra w_i que seja mais provável de ocorrer dado a etiqueta anterior e_{i-1} e a palavra atual w_i (JURAFSKY; MARTIN, 2000). Segundo Muller (2003), dentro deste contexto, esta técnica é chamada de *n-grams*, consistindo em estabelecer uma probabilidade a cada “n” palavras, normalmente com $n = 3$. Desta forma o algoritmo analisa duas palavras para prever a terceira e estabelece uma estatística na construção das frases, permitindo prever qual a próxima palavra mais provável.

2.1.2.4 POS Baseado em Transformação.

A abordagem de POS Baseado em Transformação, reúne as características das duas técnicas descritas anteriormente, utilizando tanto as regras de decisão quanto a probabilidade de Markov. Uma das técnicas mais utilizadas dentro deste contexto é o chamado Aprendizado Baseado em Transformação (ABT), que realiza uma indução de regras baseando-se em exemplos de palavras inseridas. Utilizando regras básicas, o algoritmo procura inferir a classe de POS das frases utilizadas para aprendizado, e a partir deste, construir estatísticas que podem dar origem a novas regras de construção (MULLER, 2003).

2.1.2.5 N-grams

Uma das mais importantes contribuições de técnicas estatísticas para a análise léxica e sintática, são os *n-grams*. São usados principalmente em conjunto com processadores de linguagem natural, para permitir que estes possam prever palavras em um texto (MOTE, 2002). Também chamadas de cadeias de Markov, estes modelos resolvem o problema de maneira muito simples, aproximando a probabilidade de uma palavra dado a palavra exatamente anterior. Por exemplo, ao invés de utilizar todas as palavras anteriores a próxima como $P(\text{coelho}|\text{Ontem eu vi um})$, utiliza-se apenas $P(\text{coelho}|\text{um})$ (JURAFSKY; MARTIN, 2000). Assim, podemos simplificar a sua equação em um bigrama, que é um conjunto de duas palavras sendo analisadas conforme a Equação (2):

$$P(W_n|W_{n-1}) \quad (2)$$

Para utilizar estes modelos, primeiro eles devem ser treinados com conjuntos de treinamento, para que seja possível computar a probabilidade de todos os bigramas presentes em uma sentença e quão frequentes são estes bigramas. Desta forma, o treinamento se resume na contagem dos bigramas em um determinado texto, computando todas as probabilidades (MOTE, 2002).

Com as técnicas de POS e de *n-grams* combinadas, pode-se criar os chamados *n-grams* baseados em classe, que são uma técnica que utiliza a frequência de classes de POS para ajudar a produzir uma probabilidade estimada mais inteligente de *strings* de palavras. Seu modelo básico, define a probabilidade condicional de uma palavra W_n ocorrer, dado ao produto de dois fatores: a probabilidade de uma classe C dado suas classes anteriores e a probabilidade de uma palavra dado sua classe.

N-grams baseados em classe podem ser utilizados em conjuntos específicos, como sistemas para linhas aéreas que considerem cidade e linha aérea. As classes também podem ser induzidas automaticamente, agrupando palavras de um determinado texto. Alguns pesquisadores, mostram que *n-grams* baseados em classes, podem ser úteis na diminuição da taxa de erro de palavras nos modelos de linguagem, principalmente se combinadas com base em palavras *n-grams* (JURAFSKY; MARTIN, 2000).

2.1.2.6 Árvores de Parser

Segundo Muller (2003), as técnicas de POS não são suficientes para acabar com todas as ambiguidades, onde por várias vezes se faz necessário utilizar mais um nível de processamento, pois dependendo de onde se encontra uma palavra na sentença, ela pode ter funções diferentes, por exemplo a palavra “branco” que pode ser classificada como um advérbio em algumas frases ou substantivo em outras. Desta forma, Muller (2003) sugere uma análise de construção gramatical da sentença, para que possa ser definido sua categoria sintática. Na Figura 6, tem-se a construção de uma árvore sintática para a frase “o gato caçou o rato”, representações desta forma, permitem com que a construção gramatical da sentença seja analisada de uma maneira mais específica.



Figura 7. Árvore sintática da frase “o gato caçou o rato”.

Fonte: Pereira (2010).

Após a tarefa de POS, é necessário verificar a construção gramatical das sentenças, para isso deve-se modelar a linguagem através de gramáticas livres de contexto, dividindo as frases em árvore sintagma nominal e verbal e a partir daí, verificar a qual classe se encaixa cada um dos sintagmas. Desta forma, existem duas correntes de desenvolvimento: a construção de árvores de *parser* e de *parsers* probabilísticos (JURAFSKY; MARTIN, 2000 apud MULLER, 2003). No Quadro 2 tem-se um exemplo da etiquetagem realizada na criação de árvores de *parser*.

Quadro 2 - Exemplo de etiquetagem de frase para árvore de *parser*

Etiqueta	Descrição	Palavra
PPE	Pronome Pessoal	Eu
VP	Verbo no Passado	Tropecei
PAF	Preposição e Artigo feminino	Na
SSF	Substantivo feminino no singular	Pedra

Fonte: adaptado de Muller (2003).

Árvores de *parser* são estruturas de busca através de todo o espaço de análise possível. Funcionando de forma parecida aos autômatos finitos, onde os possíveis caminhos são definidos conforme suas estruturas, entretanto o espaço de busca das árvores de *parser* são definidos por sua gramática. O objetivo destas estruturas, é encontrar todas as árvores possíveis, partindo de sua raiz e cobrindo todas as palavras do texto de entrada (JURAFSKY; MARTIN, 2000). Uma das técnicas mais difundidas para a solução do problema de busca em árvores é o algoritmo de Early (MULLER, 2003). Na Figura 8 se tem um exemplo de árvore de *parser*, criada a partir das informações do Quadro 2.

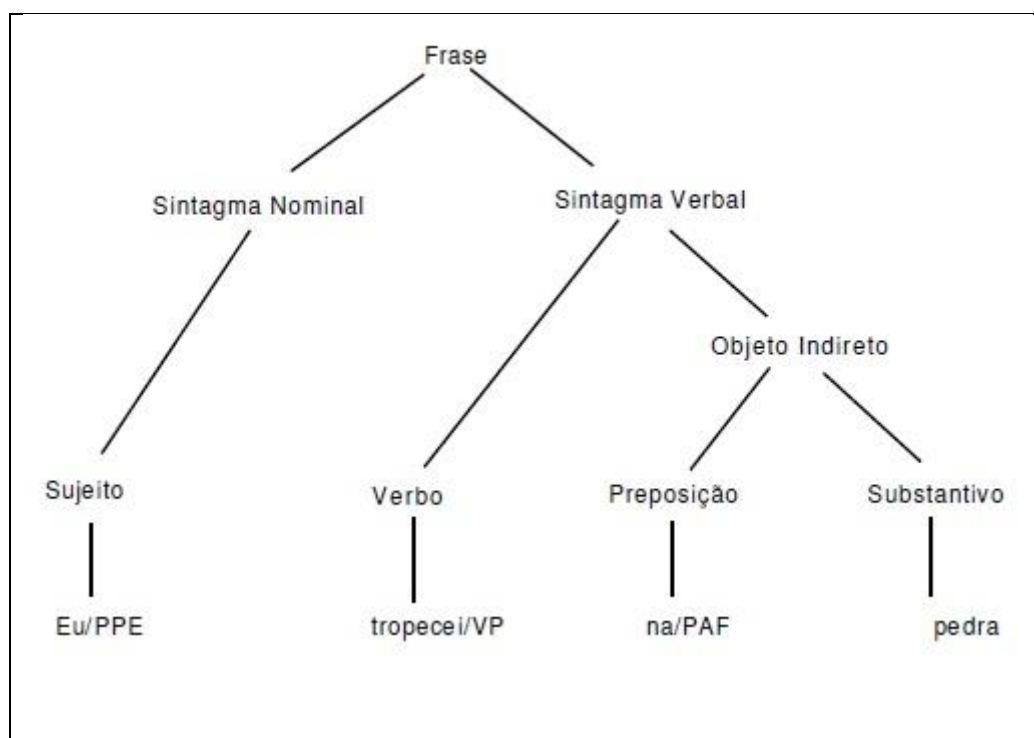


Figura 8. Exemplo de árvore de *parser*.

Fonte: adaptado de Muller (2003).

O algoritmo de Early, realiza uma busca *top-down* paralela utilizando um vetor chamado “mapa”, onde para cada posição da palavra este mapa contém uma lista de estados representando o estado parcial da árvore gerada até então (JURAFSKY; MARTIN, 2000).

Esta lista de estados possui subárvores que contêm regras gramaticais, com informações a respeito do preenchimento da árvore e da posição na subárvore de acordo com a palavra do mapa (MULLER, 2003). Ainda segundo Muller (2003), o algoritmo de Early possui três operadores: (i) o *predictor*, que alocando as folhas na árvore para estados não-terminais é responsável por criar a subárvore com base na regra gramatical; (ii) o *scanner*, que é responsável por verificar qual ramificação gerada é compatível com a palavra analisada e gerar uma nova folha para a árvore, marcando no mapa o atual estado do algoritmo; e (iii) o *completer*, utilizado quando uma regra chegou ao fim da análise, ele é responsável por fazer o retrocesso (*backtracking*), ou seja, voltar atrás para continuar analisando as demais ramificações da árvore.

Uma variação do algoritmo de Early, é o algoritmo de Early probabilístico, seguindo a mesma lógica descrita anteriormente, porém com a adição de ponderações estatísticas nas regras gramaticais. Os pesos são atribuídos com base na probabilidade de ocorrência das regras gramaticais no texto analisado, com a ajuda destes pesos é possível tratar ambiguidades, onde são indicadas as frases com a maior probabilidade de ocorrência. Desta forma, a probabilidade de uma regra R não-terminal ocorrer em uma sentença S , é a divisão da quantidade de ocorrências de R na avaliação de S pelo número de repetições do símbolo avaliado (MULLER, 2003).

2.1.3 Análise Semântica

Semântica é o estudo do significado de enunciados linguísticos. Esta análise é realizada para se estabelecer representações que captam o significado de expressões linguísticas e desenvolver algoritmos capazes de mapear enunciados linguísticos com significados apropriados. Nesta análise, é importante entender como o significado de uma expressão está relacionado com o significado das frases, palavras e morfemas que a compõem (JURAFSKY; MARTIN, 2000).

Apesar do extenso processamento realizado nas fases de análise morfológica e sintática, certas classes de palavras ainda não podem ser distinguidas, sendo impossível desta forma definir o objetivo da sentença, se fazendo necessário a adição de anexos semânticos aos *parsers*, por exemplo. Esta análise pode ser realizada em conjunto com as árvores de *parsers*, a medida que esta vai sendo gerada, pode-se incluir os anexos semânticos. Além deste método, pode-se destacar também gramáticas semânticas e gramáticas baseadas em caso

(MULLER, 2003). Na Figura 9, pode-se observar a construção da mesma árvore mostrada na Figura 8, porém com o anexo semântico a frase

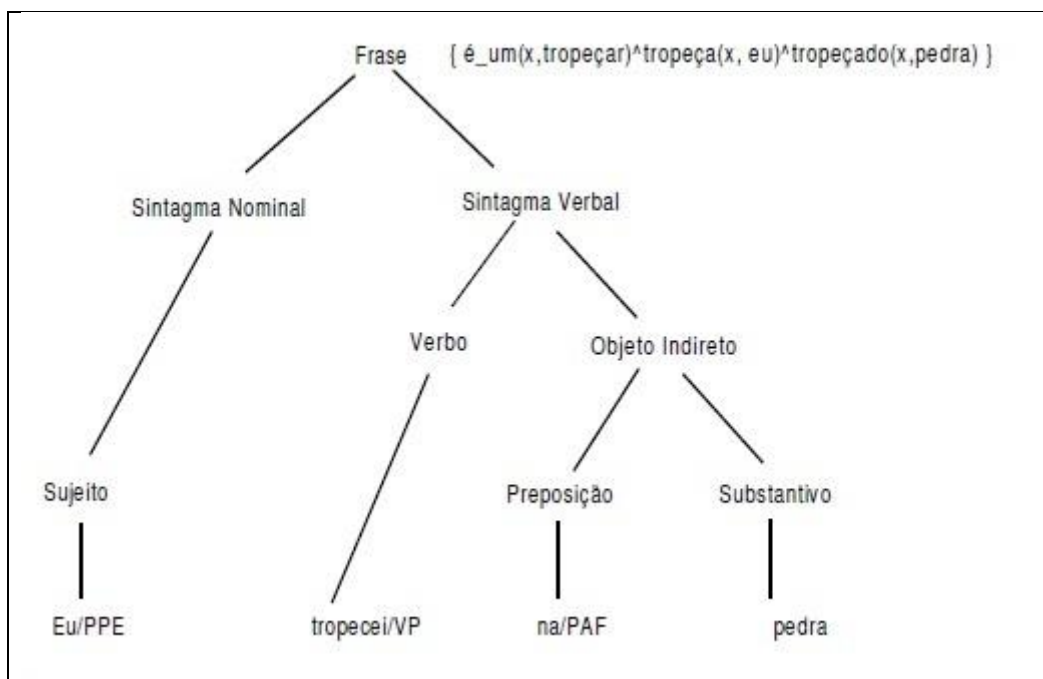


Figura 9. Exemplo de árvore de *parser* com anexo semântico.

Fonte: adaptado de Muller (2003).

2.1.3.1 Gramática Semântica

Originalmente desenvolvida para sistemas de diálogos baseados em texto com o objetivo de responder perguntas, gramáticas semânticas foram adaptadas para resolver as necessidades de análise da linguagem. Nestes casos, as regras que constituem a gramática são pensadas especificamente para atender as entidades e relacionamentos do domínio da análise composicional, ou seja, são construídas para que os componentes da chave semântica possam ocorrer juntos com regras únicas e as regras são feitas para atender as necessidades da análise semântica.

Segundo Muller (2003), a construção das gramáticas semânticas baseia-se na criação de variáveis que devem ser preenchidas de acordo com a afirmação ou questionamento dos usuários de um sistema. Exemplificando, Muller (2003) traz:

Quadro 3 - Exemplo de gramática semântica.

Variáveis	Expressão
QUESTÃO	Quem tropeçou ONDE
RESP_QUEST	QUEM tropeçou ONDE
QUEM	Eu
ONDE	PREPOSIÇÃO LUGAR
PREPOSIÇÃO	Na
LUGAR	Pedra

Fonte: adaptado de Muller (2003).

No Quadro 3, para o exemplo da árvore de *parser* “Quem tropeçou na pedra?”, a resposta seria “eu tropeçou na pedra”, associando a expressão “na pedra” com a variável ONDE e o pronome “eu” com a variável QUEM.

Dentre as motivações para o uso de gramáticas semânticas, pode-se destacar a necessidade de lidar com anáforas e elipses. Elas conseguem resolver fenômenos de repetição e omissões de palavras, pois permitem que os analisadores façam previsões altamente específicas sobre a próxima entrada de palavra com base nas categorias. Devido a esta possibilidade, as referências anafóricas e os elementos faltantes podem ser associados a categorias semânticas específicas.

2.1.3.2 Gramáticas Baseadas em Caso

Casos representam as relações semânticas expressadas sintaticamente por verbos e seus argumentos. Verbos expressam ações, e seus argumentos expressam as circunstâncias que envolvem essas ações. Pesquisadores que desenvolvem sistemas de PLN baseados em casos, empregam bastante esforço no desenvolvimento de conjuntos de casos. Os sistemas baseados em caso destes estudos, explicitam o ponto de partida para aplicações de extração de informação e aquisição de conhecimento de textos (BARKER, 1998).

Segundo Muller (2003), são utilizadas gramáticas semânticas não-terminais para se formar os padrões desejados para estas análises. Se uma sentença encaixa na construção padrão, ela poderá ser reconhecida dentro de seu contexto, que normalmente é estruturado na forma de casos. Constrói-se as gramáticas semânticas, que identificam os elementos na sentença e indicam o caso que melhor se aplica aos elementos. Muller (2003), exemplifica utilizando o seguinte caso:

Quadro 4 - Exemplo de caso.

Variáveis	Expressão
< msg >	<? Artigo cabeça – msg * - msg >
< caso – msg >	< % de Pessoa >
< caso – msg >	< %sobre Assunto >
< caso – msg >	< %desde Data >

Fonte: adaptado de Muller (2003).

No Quadro 4, o símbolo “?” indica que um item é opcional, “*” que um item pode se repetir e “%” significa que qualquer palavra da mesma classe indicada pode ser utilizada. A palavra “artigo” indica que pode haver um artigo antecedendo o início da sentença; “cabeça-msg” é a variável que carrega informações da origem da mensagem. A palavra-chave “caso-msg” refere-se ao caso que irá tratar aquela parte da sentença. Por fim, as variáveis “Pessoa”, “Assunto” e “Data” recebem os dados sobre a mensagem.

A criação dos casos para a análise semântica antecipa um processamento realizado normalmente na fase de análise pragmática, utilizado para descrever contextos de sentenças. Na construção dos casos, são utilizados quadros, para especificar os elementos semânticos que compõem a sentença sendo analisada. Desta forma, os casos são estruturas que contém quadros e são utilizados em determinado contexto para identificar problemas e suas soluções. Eles são utilizados normalmente para resolver problemas de ambiguidade não identificados na análise baseada em regras. Alguns pesquisadores sugerem que seja feito a análise pragmática juntamente com a semântica utilizando-se os casos, estes argumentam que a abordagem do estudo das relações existentes na sentença de um caso, dependendo da relação, é definido a função da frase dentro do caso, desta forma obtém-se ao mesmo tempo o aspecto semântico (significado) e pragmático (contexto) (MULLER, 2003).

2.1.4 Análise Pragmática

A análise pragmática é o estudo da relação entre linguagem e contexto, como a identidade da pessoa e objetos na sentença. Desta forma, a pragmática estuda o meio como a linguagem é utilizada para referenciar pessoas e coisas. O contexto de uso inclui o contexto do discurso, assim estuda-se também como o discurso é estruturado e como o ouvinte interpreta a conversação (JURAFSKY; MARTIN, 2000). Este tipo de análise, não foca apenas na estrutura de uma frase, mas sim em buscar nas demais sentenças a compreensão de contexto da frase em análise (MULLER, 2003).

Segundo Muller (2003), geralmente não existem estruturas pré-definidas para representar adequadamente o problema de referências pronominais, por exemplo o uso do pronome “seu” e “o” na sentença “João pegou a rosa. Ao pegá-la, seu espinho o espetou”, bem como coerência textual e análise de discurso. As estruturas mais utilizadas na análise pragmática, são os casos, descritos anteriormente, que permitem o entendimento do contexto das sentenças. O problema de referência, que se apresenta como no exemplo acima, em PLN pode ser chamado de expressão de referência e a entidade a que se refere, chamada de referente. Desta forma, no exemplo “João pegou a rosa. Ao pegá-la, seu espinho o espetou”, a sentença pode ser considerada uma expressão de referência e “João” e a “rosa” seus referentes (JURAFSKY; MARTIN, 2000). Para se tratar esses casos de referências pronominais, Jurafsky e Martin (2000) descrevem um algoritmo de Lappin e Leass (1994) que utiliza um sistema de pesos simples para estabelecer prioridades na sintaxe. Existem dois tipos de operação efetuadas pelo algoritmo: (i) atualização do modelo de discurso; e (ii) resolução de pronome. Toda vez que uma expressão de referência é identificada, é feito uma representação para ela no modelo de discurso e atribuído a ela um grau de saliência, que é calculado com a soma dos pesos atribuídos de acordo com um conjunto de fatores de saliência.

Tabela 1 - Fatores de saliência

Fator	Valor
Recência de Sentença	100
Ênfase no Sujeito	80
Ênfase no Predicado Nominal	70
Ênfase no Objeto	50
Ênfase no Objeto indireto	40
Ênfase Adverbial	50
Ênfase no Substantivo	80

Fonte: adaptado de Jurafsky e Marin (2000).

Os pesos que cada fator possui no modelo de discurso diminuem pela metade a cada nova sentença processada. A recência de sentença, recebendo peso cem, e sendo diminuído pela metade a cada iteração, permite que o algoritmo dê preferência a frases mais recentes (a sentença atual tende a ter um peso maior do que a sentença anterior, e assim sucessivamente). Os próximos cinco pesos da Tabela 1 estabelecem um papel de preferência gramatical, assim temos a seguinte precedência: sujeito > predicado nominal > objeto > objeto indireto > advérbios.

Por último, o fator de substantivo é utilizado para penalizar os referentes que fazem parte de sentenças nominais maiores, promovendo o peso dos referentes de frases menores.

Cada um desses fatores contribui para a relevância dos referentes com base nas propriedades nominais, sendo assim pode ser que várias expressões de referência se refiram ao mesmo referente das sentenças anteriores, cada uma com um fator de saliência diferente. Lappin e Leass (1994), sugerem associar cada referente a uma classe de equivalência contendo todas as expressões de referência que se referem a ela. O peso da saliência para cada referente é a soma dos pesos de cada fator levando em consideração sua classe de equivalência. Portanto, se considerarmos o escopo de um fator de saliência uma sentença, e sendo um referente mencionado na sentença atual, bem como na anterior, o peso da referência será considerado individualmente para cada um, fazendo com que referentes múltiplos nas sentenças anteriores tenham um peso maior. Uma vez atualizado o modelo de discurso, passa-se a considerar o processo de resolver qualquer pronome que exista dentro de uma sentença. Para isso, utiliza-se mais dois fatores com pesos de saliência, conforme observa-se na Tabela 2. O paralelismo gramatical, usado para tratar a regra gramatical entre pronome e seu referente. E a catáfora, para desprezar referências catafóricas (JURAFSKY; MARTIN, 2000). Referências catafóricas, são referências feitas nas sentenças para indicar termos que ainda estão por vir. Quando a palavra de coesão aparece antes do termo pressuposto. Por exemplo, “Os políticos poderiam fazer uma coisa: renunciar à imunidade”, a expressão “uma coisa” faz referência a parte da sentença “renunciar a imunidade”, onde ocorre a chamada referência catafórica (LIMA, 2011).

Tabela 2 - Complemento de fatores de saliência

Fator	Valor
Paralelismo Gramatical	35
Catáfora	-175

Fonte: adaptado de Jurafsky e Marin (2000).

Com estes pesos atribuídos pode-se especificar o algoritmo de resolução de pronome: (i) coletar os referentes; (ii) remover referentes que não se encaixem em número ou gênero com o pronome; (iii) remover referentes que não tenham coerência sintática; (iv) somar o valor total de saliência; e (v) selecionar o referente com o maior valor de saliência. Este algoritmo, quando combinado com filtros os quais o autor não descreve, conseguiu alcançar um nível de acerto de 86% quando aplicado a dados de teste desconhecidos de mesmo gênero (JURAFSKY; MARTIN, 2000). Segundo Muller (2003), existem outros algoritmos a respeito de resolução de referências pronominais, entretanto a maioria deles depende de uma estrutura sintática previamente estabelecida e/ou pouco flexível às diferenças de construção das

sentenças. Devido a esta dificuldade, tem sido cada vez mais comum o uso de estruturas de preenchimento como os casos, estudados anteriormente, para se efetuar análise pragmática.

2.2 ANÁLISE DE SENTIMENTO

Análise de Sentimento (AS), também chamada de Mineração de Opinião (MO) é o estudo computacional de sentimentos, opiniões, atitudes e emoções expressadas por pessoas a respeito de uma entidade, que pode representar um indivíduo, evento ou tópico. Destes tópicos estudados em AS, destacam-se opiniões de usuários na internet. AS pode ser considerada um processo de classificação e existem três níveis principais. O nível de documento, que visa classificar a opinião de um documento expressando um sentimento positivo ou negativo. Considera todo o documento como base de informação para a tarefa. O nível de sentença, que visa classificar o sentimento expressado em cada sentença da entidade, primeiro identifica se a sentença é objetiva ou subjetiva, e caso seja subjetiva, esta tarefa determina se a sentença expressa um sentimento positivo ou negativo. Finalmente, o nível de aspecto, visa classificar o sentimento que diz respeito aos aspectos específicos de cada entidade. Primeiro, identifica-se a entidade em si e seus aspectos e pode se extrair diferentes opiniões para diferentes aspectos de uma entidade (MEDHAT, HASSAN, KORASHY, 2014).

A tarefa de AS a nível de documento, tem por objetivo classificar se um documento inteiro de opinião expressa um sentimento positivo ou negativo. Por exemplo, em um documento que possui o conjunto de revisões sobre determinado produto, o sistema irá determinar se estas revisões expressam uma opinião positiva ou negativa a respeito do produto. Esta tarefa é chamada de classificação de sentimento a nível de documento, e pressupõe que cada documento expressa opiniões sobre uma única entidade (um único produto, por exemplo). Não é aplicável a documentos que avaliam ou comparam várias entidades. Ao se processar a tarefa de AS a nível de sentença, deve-se analisar se cada sentença expressa uma opinião positiva, negativa ou neutra (não expressa opinião). O objetivo deste nível de análise é realizar uma classificação subjetiva, distinguindo frases que expressam fatos (sentenças objetivas) de frases que expressam pontos de vista e opiniões. É importante destacar que a subjetividade não é equivalente a sentimento ou opinião, pois existem sentenças objetivas que podem insinuar opiniões, como por exemplo “Comprei um carro e em menos de um mês precisei trocar todas as lâmpadas dos faróis”, a sentença irá estabelecer um fato, porém também pode implicar em uma opinião negativa sobre a entidade carro (LIU, 2012).

Ainda segundo Liu (2012), as análises a nível de documento e sentença, não são suficientes para extrair a opinião exata expressada pelas pessoas. Para que se faça uma análise mais refinada, utiliza-se a tarefa de AS a nível de entidade e aspecto. Ao invés de analisar a construção da linguagem (documentos, sentenças ou frases), a análise a nível de aspecto processa diretamente a opinião em si. Esta tarefa parte do princípio que uma opinião expressa um sentimento, positivo ou negativo, e um alvo (de opinião). Uma opinião sem um alvo tem um uso limitado para a análise a nível de aspecto. Liu (20012), exemplifica com a frase “Apesar do serviço não ser ótimo, eu ainda amo esse restaurante”, onde apesar do tom da sentença ser, de maneira geral, positivo, não se pode afirmar que a frase expressa uma opinião completamente positiva. A frase é explicitamente positiva a respeito do alvo “restaurante” e implicitamente negativa a respeito do alvo “serviço”, nesses casos os alvos de opiniões podem ser descritos pelas entidades e/ou seus aspectos diferentes.

Assim, com este nível de análise, pode-se criar um sumário de opiniões sobre entidades e seus aspectos, transformando textos em dados estruturados, que podem ser utilizados para várias análises quantitativas e qualitativas. Além dos níveis em que se pode realizar as tarefas de AS, existem diferentes técnicas que podem ser aplicadas para se obter resultados. Medhat, Hassan e Korashy (2014), categorizaram mais de cinquenta artigos recentes de acordo com a técnica utilizada, dando uma visão panorâmica da área, permitindo utilizar sua classificação para nortear os estudos, definindo quais técnicas de AS tem maior importância de estudo recentemente. Na Figura 10, apresenta-se uma adaptação das técnicas identificadas no trabalho de Medhat, Hassan e Korashy (2014), onde estão representadas algumas das técnicas utilizadas em AS.

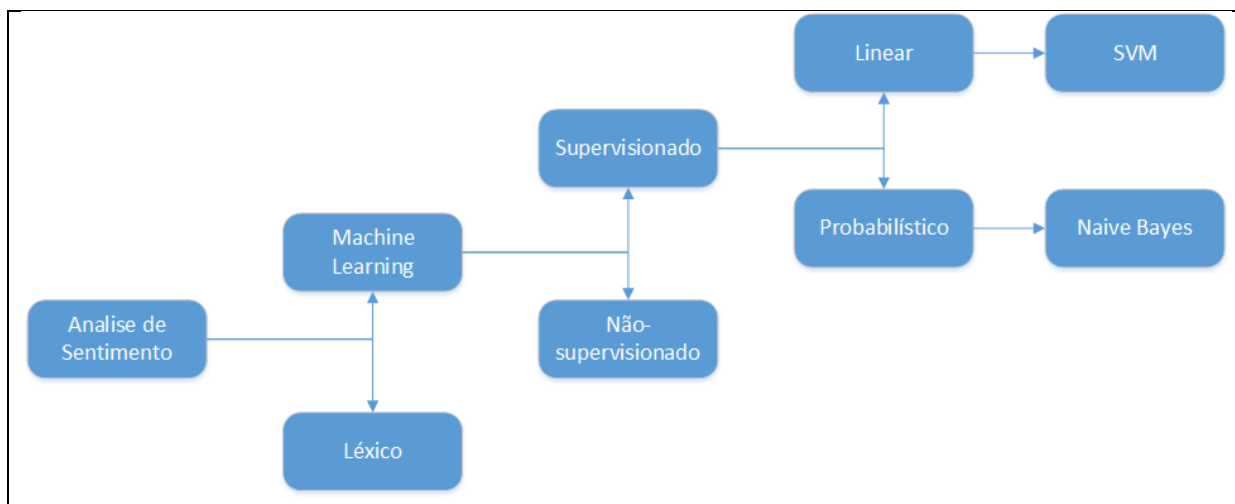


Figura 10. Técnicas utilizadas para a tarefa de AS.

Fonte: adaptado de Medhat, Hassan e Korashy (2014).

Entre as formas mais utilizadas para se realizar a tarefa de AS, pode-se destacar técnicas de Machine Learning, que tratam o problema de análise de sentimento como uma classificação de texto, fazendo uso de características de linguagem e análises sintáticas, podendo ser supervisionado ou não. Destacam-se também os métodos Léxicos, que podem utilizar um conjunto de documentos para realizar uma análise estatística ou semântica, ou com um dicionário de palavras, que utiliza um conjunto de palavras inseridas manualmente com opinião conhecida.

2.2.1 Machine Learning

Machine Learning (ML) é uma das abordagens utilizadas para solucionar o problema de classificação que a AS propõe, normalmente descrito em dois passos gerais: (i) fazer com que o modelo aprenda a partir de um conjunto de documentos, chamado de corpus ou base de treinamento; e (ii) classificar os dados que se deseja analisar a partir do modelo treinado (TSYTSARAU; PALPANAS, 2012).

Para resolver o problema, a abordagem de ML trata a AS como um problema de classificação de texto comum, que utiliza recursos sintáticos e linguísticos. Com um conjunto de registros de treinamento (supervisionado) ou não, onde normalmente cada um dos registros é rotulado com uma classe, este modelo está diretamente relacionado às características que estão submetidas aos registros, rotulados com uma classe específica. Os métodos de ML podem ser utilizados geralmente sob duas perspectivas: (i) supervisionado, utilizando conjuntos de treinamento para atribuir rótulos de sentimentos e algoritmos complexos para

realizar a classificação; ou (ii) não supervisionado, que ignora as dificuldades de se obter um conjunto treinado com rótulos de sentimento pré-estabelecidos e propõe métodos que utilizam os próprios documentos a serem analisados para solucionar o problema. Nesses casos, normalmente os documentos são divididos em sentenças e classificados com o uso de listas de palavras-chave para cada categoria, medindo sua similaridade (MEDHAT, HASSAN, KORASHY, 2014).

2.2.1.1 Supervisionado

Métodos de ML supervisionados utilizam um conjunto de treinamento. A performance destes métodos, está diretamente ligada a qualidade e a quantidade dos dados de treinamento, independentemente da escolha de algoritmo (TSYTSARAU; PALPANAS, 2012). Existe uma variedade de algoritmos supervisionados de ML, dentre as técnicas utilizadas atualmente, estão: (i) classificadores lineares, que geralmente utilizam vetores de coeficientes lineares para realizar a busca e a classificação; e (ii) classificadores probabilísticos, que utilizam um gerador de amostras de probabilidade para termos em particular (MEDHAT, HASSAN, KORASHY, 2014).

Classificadores Lineares

Os classificadores lineares, geralmente utilizam vetores de coeficientes lineares para realizar uma busca e efetuar a classificação. O SVM (*Support Vector Machines* – Máquina de Suporte de Vetores) é um método de ML linear supervisionado, que funciona com base na construção de um hiperplano para separar classes, de maneira que se estabeleça a maior margem de separação possível entre elas. Esta margem é definida pela soma das distâncias do hiperplano para pontos de dados mais próximos de cada classe e estes pontos são chamados de Vetores de Suporte (do inglês *Support Vectors*) (BYUN; LEE, 2002 apud ROMASZEWSKI et al., 2013). O princípio do SVM é determinar separadores lineares que melhor dividam as diferentes classes no espaço de busca. Na Figura 11, pode-se observar que existem duas classes “+” e “o”, e três margens para o hiperplano, representadas por A B e C. Onde a margem A oferece a melhor separação entre as duas classes, pois oferece a maior distância de qualquer ponto de classes, desta forma representando a maior margem de separação (MEDHAT, HASSAN, KORASHY, 2014).

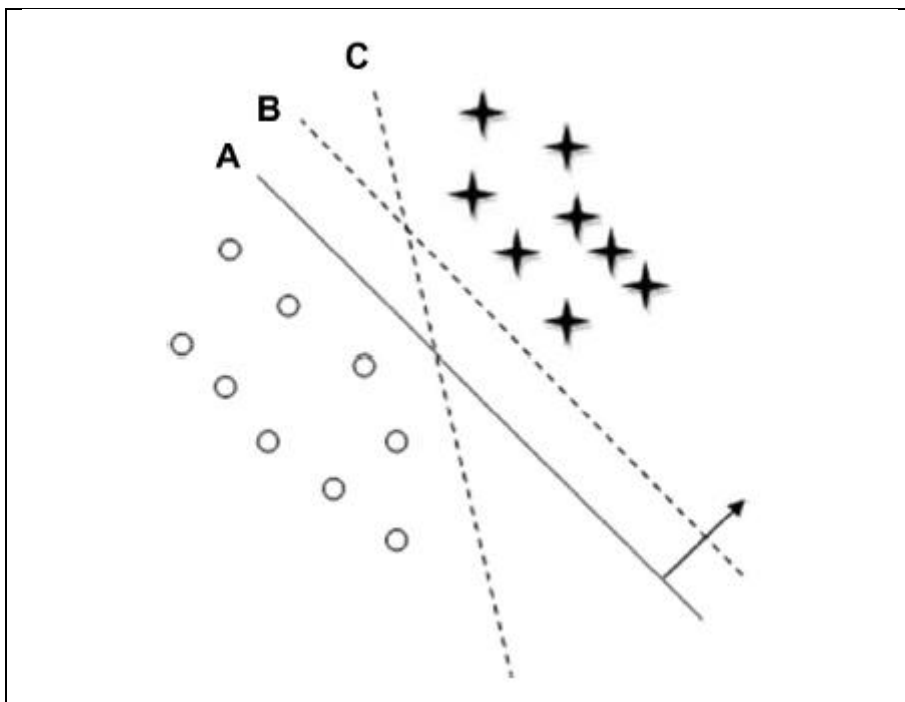


Figura 11. Representação de um hiperplano da técnica de SVM.

Fonte: Medhat, Hassan e Korashy (2014).

Li e Li (2013), descrevem um sistema de mineração de opinião em microblogs, onde fazem uso, em seu módulo de AS, da técnica de SVM. Neste módulo o modelo SVM é treinado e usado para classificar a polaridade da opinião através de três etapas: (i) selecionar as características nos textos; (ii) criar um conjunto de dados de treinamento devidamente rotulado com suas classes de palavras; (iii) encontrar a melhor combinação de parâmetros e configurações de entrada do modelo para efetuar a classificação. Como mensagens de microblogs são curtas, a chance de repetição de características na mesma mensagem é pequena. Desta forma, todas as características encontradas são contadas como um valor binário $\{0, 1\}$, onde 1 representa o aparecimento de uma característica e 0 a ausência da mesma. O sistema utiliza um kernel RBF (Radial Basis Function) e pratica uma busca de grid para encontrar o melhor par de parâmetro. Com o SVM treinado, a polaridade da opinião de $polaridade_o \in \{+1, -1\}$ pode ser extraída. Caso seja positiva (+1) ou negativa (-1) (LI; LI, 2013). Um kernel RBF é definido como uma matriz utilizada para separar os vetores de treinamento em um espaço mapeado (LIN; LIN, 2003). A busca de grid é uma das técnicas de buscas mais utilizadas para a otimização de hiperparâmetros, ela testa exaustivamente os pares de parâmetros necessários para o kernel RBF, e com um modelo de validação cruzada é capaz de determinar a melhor combinação de par de parâmetro (HSU; CHANG; LIN, 2016). Por fim, calcula-se a pontuação semântica através da Equação (3):

$$PS_{o,t} = polaridade_o \cdot SO_{o,t}, \text{ onde } PS_{o,t} \in [-1, 1] \quad (3)$$

Na Equação (3), $PS_{o,t}$ é a pontuação semântica de uma opinião o a respeito de um tópico t . A polaridade de uma opinião é multiplicada pelo valor de subjetividade da opinião representado por $SO_{o,t}$. O valor de subjetividade é usado para diminuir a incapacidade do SVM de filtrar opiniões neutras. Para se chegar ao valor de subjetividade é criado um conjunto de palavras subjetivas, emocionais e sentimentais. Então é realizada a soma de todas as palavras subjetivas no texto referentes ao tópico e comparado ao conjunto criado para determinar a relevância da subjetividade (LI; LI, 2013).

Para o treinamento e avaliação do modelo, foi criado um conjunto de treinamento automaticamente baseado em *emoticons*. A preparação do conjunto de treinamento é feita com 11.929 postagens que contém os *emoticons* “:)” rotulados com a classe positiva “+1” e “:(” rotulados com a classe negativa “-1”, caso a postagem tenha os dois *emoticons* é descartada. Após o processo de rotulagem obteve-se 7.510 postagens positivas, 3.947 negativas e 236 postagens foram descartadas. As postagens positivas e negativas foram separadas em cinco grupos aleatórios, onde os quatro primeiros foram utilizados para treinamento, somando 9.165 postagens. Os demais foram utilizados para teste. Antes de serem utilizados, os conjuntos foram pré-processados afim de remover tópicos específicos para que o método não classifique o sentimento por um tópico em particular. Os números encontrados nas postagens foram substituídos por um *token* chamado de “NUMBER_OR_VERSION” e foi adicionado o prefixo “NÃO_” em cada palavra precedida de uma palavra que expresse negatividade na sentença. Por último, extraiu-se diferentes conjuntos de características e se avaliou suas precisões. (LI; LI, 2013).

Conjuntos de apenas uma palavra, obtiveram o melhor nível de precisão. Isso pode ser explicado porque textos de microblogs são mais informais e possuem “termos inventados” ou gírias usadas para expressar sentimentos, este fato afeta o conjunto de palavras subjetivas, uma vez que este foi criado com base em palavras do dicionário. O algoritmo SVM obteve uma precisão máxima de 90,4%, quase 20% a mais que a técnica probabilista Naive Bayes, utilizada para comparação, que obteve uma precisão máxima de 71,7% comparada ao mesmo conjunto de dados e treinamento (LI; LI, 2013).

Classificadores Probabilísticos

Classificadores probabilísticos utilizam modelos de mistura para realizar a classificação, estes modelos assumem que cada classe é um componente da mistura e cada componente de mistura é um modelo generativo, ou seja, ele gera a probabilidade amostral de um termo específico para o componente. Estes classificadores também são chamados de classificadores generativos e o mais utilizado é o Classificador Naive Bayes (do inglês Naïve Bayes Classifier) (MEDHAT, HASSAN, KORASHY, 2014).

O Classificador Naive Bayes (CNB) computa a probabilidade de uma classe ocorrer, baseado na distribuição das palavras no documento. Esta técnica se baseia no teorema de Bayes para prever a probabilidade de um conjunto de características pertencerem a um determinado rótulo, pode ser descrito através da Equação (4):

$$P(l|c) = \frac{P(l) * P(c|l)}{P(c)} \quad (4)$$

Onde $P(l)$ é a probabilidade prévia de um rótulo ou recurso aleatório de definir um rótulo. $P(c|l)$ é a probabilidade prévia de um conjunto de características serem classificadas com um rótulo. $P(c)$ é a probabilidade prévia de que um conjunto de características ocorra (MEDHAT, HASSAN, KORASHY, 2014).

Kang e Yoo (2011), descrevem um sistema que utiliza a técnica probabilística do CNB, porém com um algoritmo melhorado, chamado INB-1 combinado com a técnica de SVM para realizar análise de sentimento em revisões de clientes de restaurante. O trabalho coletou opiniões de clientes a respeito de restaurantes e gerou o que é chamado de “senti-lexicon”, que pode ser traduzido basicamente como um dicionário de palavras que expressam sentimentos. O conjunto léxico criado inclui padrão de bigramas, com palavras em dupla, desta forma pode identificar, tanto padrões de uma palavra como “bom”, “ruim” e “delicioso”, como também de duas “muito ruim”, “nada bom” ou “pouco bom”. Utilizando essa base léxica o sistema combina os algoritmos e características das técnicas de CNB e SVM a fim de melhorar a qualidade dos resultados.

O algoritmo de SVM é utilizado para classificar o texto como positivo ou negativo, separando linearmente os diferentes tipos de vetores inseridos em um hiperplano. A função de decisão utilizada para determinar o hiperplano ótimo é visualizada na Equação (5):

$$f(x) = \sum_{i=1}^n a_i y_i(x_i, x) + b \quad (5)$$

Onde $f(x)$ é a função de decisão que determina o hiperplano ótimo dado um vetor x inserido. O símbolo a é um vetor de suporte utilizado para determinar se o parâmetro é positivo ou negativo e y é uma classe que pode conter o valor 1, caso seja positivo, ou -1 caso seja negativo. A letra x representa um vetor inserido e b é a distância da origem até o hiperplano. Ao se realizar a classificação utilizando um conjunto léxico combinado ao SVM, é configurado um padrão geral, indicando a quantidade de palavras avulsas e em pares que serão identificadas, e um padrão é incluído no conjunto léxico. Além disso atribui-se pesos aos padrões, onde o padrão geral é configurado com peso 0,5 e o padrão léxico como 1, assim atribui-se um peso maior ao valor do padrão de sentimento incluído dentro do conjunto léxico (KANG; YOO, 2011).

O segundo algoritmo descrito e utilizado no sistema para realizar a classificação e comparar com o algoritmo melhorado, é o CNB. Ele realiza a análise de sentimento baseado na probabilidade condicional conforme as Equações (6), (7) e (8):

$$Class(d_i) = \arg \max P(c_j) \prod_{i=1}^d P(p_i | c_j) \quad (6)$$

$$P(p_i | c_j) = \frac{|c_j| + 1}{|V| + \sum_{i=q}^{|d|} \sum_{k=1}^{|c|} f(p_{ik})} \quad (7)$$

$$P(c_j) = \frac{|c_j|}{\sum_{k=1}^{|C|} |c_k|} \quad (8)$$

Na Equação (6) tem-se $Class(d_i)$ como a função que determina a classe, podendo ser positiva ou negativa, do documento d_i . A probabilidade de c_j é calculada através de $P(c_j)$ e a probabilidade do padrão p_i pertencer a classe c_j é calculado a partir de $P(p_i | c_j)$. Na Equação (7), o símbolo $|c_j|$ representa a contagem de padrões incluídos na classe c e o número 1 é adicionado na equação para que a contagem não seja zero. Em $|V|$ obtém-se a contagem total

de padrões de todo o documento e $\sum_{i=q}^{|d|} \sum_{k=1}^{|c|} f(p_{ik})$ calcula a frequência do padrão de todo o documento. Na Equação (8) o símbolo $|c_j|$ também representa a contagem de padrões incluídos na classe c e $\sum_{k=1}^{|c|} |c_k|$ calcula a contagem total de padrões no documento inteiro. Ao contrário da técnica de SVM, o algoritmo CNB não necessita de um vetor como parâmetro para realizar a AS, ele a realiza baseando-se no valor de cada característica (KANG; YOO, 2011).

O algoritmo melhorado proposto por Kang e Yoo (2011), chamado de INB-1 é definido através das Equações (9) e (10):

$$Class(d_i) = \arg \max R_1(p_{ij}) P(c_j) \prod_{i=1}^d P(p_i|c_j) \quad (9)$$

$$R_1(p_{ij}) = \frac{\sum_{p_{ij} \in L_j} C(p_{ij})}{\sum_{p_{ij} \in L} C(p_{ij})} \quad (10)$$

A Equação (9) possui os mesmos valores da Equação (6), com a adição do termo $R_1(p_{ij})$, que implica na razão do padrão p_{ij} aplicada a classe j e calculada a partir do conjunto léxico com a probabilidade de j . Esta razão é multiplicada pela Equação (6) após as probabilidades positivas e negativas serem também calculadas. Na Equação (10), L refere-se ao conjunto de palavras léxico e $|L|$ ao número total de padrões do conjunto, $\sum_{p_{ij} \in L_j} C(p_{ij})$ conta todos os padrões que estão inclusos na classe j e o divide por $\sum_{p_{ij} \in L} C(p_{ij})$ que representa o número de padrões p_i que estão inclusos na totalidade da classe. $C(p_{ij})$ representa o número de padrões contados. Desta forma, assumindo que um documento d possui padrões positivos p_1, p_2, p_3 em seu conjunto léxico e p_4, p_5 são padrões negativos. Pode-se dizer que a razão de padrão positivo $R_{1(positivo)}$, é $3/5$ e a razão de padrão negativa $R_{1(negativo)}$ é $2/5$ no conjunto documento $d = \{p_1, p_2, p_3, p_4, p_5\}$.

Após o treinamento e classificação utilizando o algoritmo melhorado INB-1, o sistema conseguiu atingir um nível de precisão de 80% com uma diferença de 2% entre as classes (positivas e negativas), em dois casos de testes, obteve-se uma precisão de 81,3% com uma diferença entre as classes de 1,4%. Dobrando a quantidade de palavras no conjunto léxico, de

11 para 22, obteve-se uma melhora para 81,4% com uma diferença entre as classes de 5,2%. O algoritmo melhorado INB-1 apresentou uma melhora na precisão entre 1,9% e 5,6% se comparado ao algoritmo de CNB normal.

2.2.1.2 Não-Supervisionado

Aprender quando não existe nenhuma informação sobre o resultado correto é chamado de ML não-supervisionado. O aprendizado não-supervisionado pode aprender relacionamentos entre suas percepções usando métodos de aprendizagem supervisionados, ou seja, pode aprender a prever suas percepções, dado suas percepções anteriores, mas não pode aprender o que fazer sem uma função de utilidade (RUSSEL; NORVIG, 1995). As técnicas empregadas nestas abordagens, partem do princípio que não se conhece a classe que pertencem os exemplos e desta forma tentam encontrar nos próprios dados, similaridades ou diferenças que possam agrupar os exemplos que possivelmente pertençam as mesmas classes ou descartar os exemplos de classes distintas respectivamente (RUSSEL; NORVIG, 2002 apud STANGE, 2011).

Ko e Seo (2000), propõe um sistema para realizar a categorização automática de textos por meio de um método não-supervisionado. O sistema proposto é desenvolvido em três módulos: (i) um módulo de pré-processamento; (ii) módulo para criar um conjunto de sentenças de treinamento; e (iii) um módulo para extrair características e classificar documentos. O modelo descrito é apresentado na Figura 12.

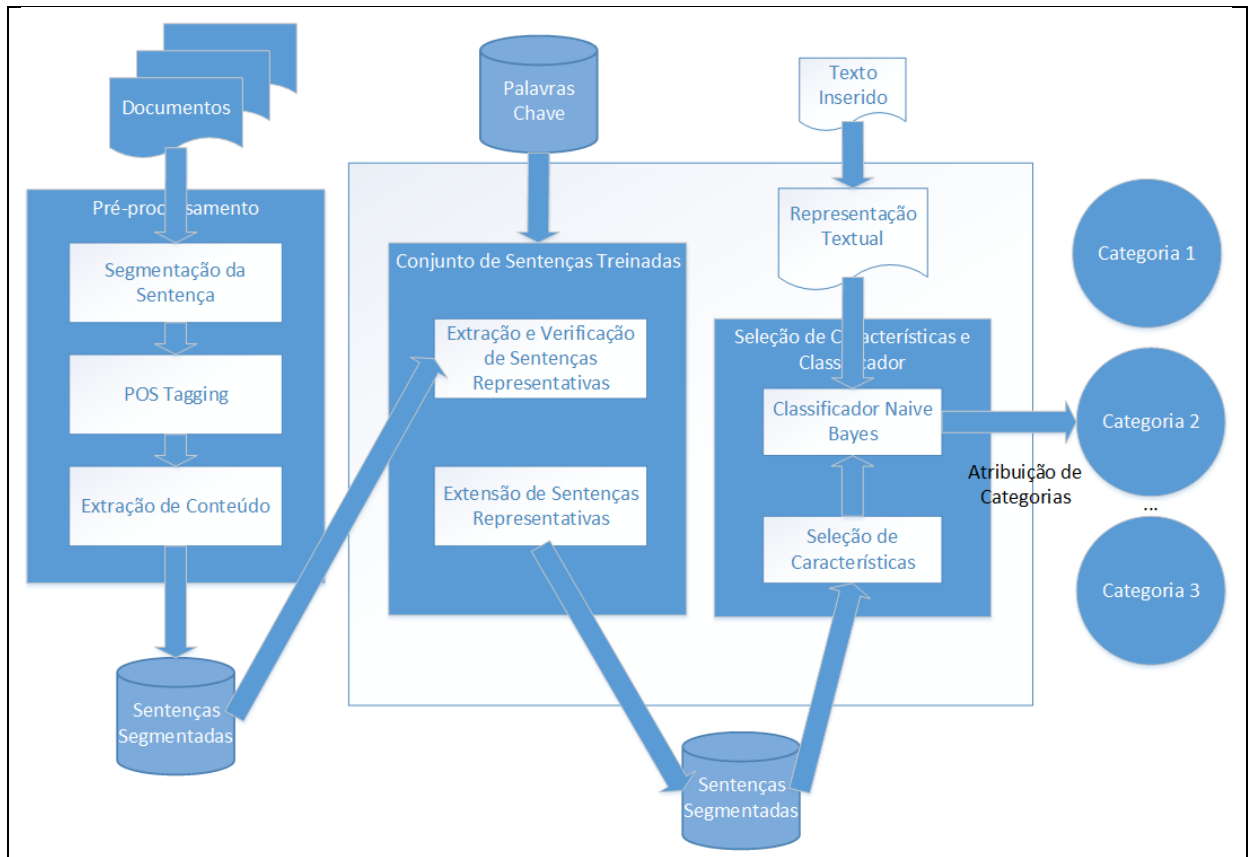


Figura 12. Modelo do sistema de análise de Ko e Seo.

Fonte: adaptado de Ko e Seo (2000).

Como o sistema proposto não utiliza conjuntos de treinamento, o conjunto de sentenças de treinamento para cada classe deve ser criado, para tal foi criado manualmente palavras-chave para cada categoria, contendo características especiais para cada classe. A escolha destas palavras-chave, foi feita com base nos nomes das categorias e seus sinônimos e incluso várias outras palavras com o significado definido de cada categoria, somando um total de 141 palavras-chave para 47 categorias. As sentenças que contém palavras-chave de categoria são escolhidas para serem a representação inicial das frases (KO; SEO, 2000).

Segundo Ko e Seo (2000), mesmo pré-definindo palavras-chave para representar as sentenças, corre-se o risco de cometer erros, pois existem sentenças que não possuem a característica especial da categoria atribuída, mesmo contendo a palavra-chave de categoria. Para remover estes erros, deve-se ranquear as sentenças representativas computando o peso de suas palavras. Estes pesos são computados através da Equação (11):

$$w_{ij} = FT_{ij} \cdot (\log(M) - \log(FC_j)) \quad (11)$$

Onde w_{ij} representa o peso calculado da palavra i para a respectiva categoria j . FT_{ij} é a frequência com que os termos aparecem para a palavra i na categoria j . Esta primeira parte é multiplicada pela frequência inversa de categoria dada por $(\log(M) - \log(FC_j))$, onde FC_j é a frequência de categoria contida na palavra e M o número total de categorias. Com o peso das palavras computado, é possível calcular o peso da sentença através da Equação (12):

$$W_{ij} = \frac{w_{1j} + w_{2j} + \dots + w_{Nj}}{N} \quad (12)$$

Onde N é o número total de palavras na sentença. Por fim, as sentenças representativas de cada categoria são organizadas em ordem decrescente de peso calculado na Equação (10), as 70% melhores sentenças representativas são selecionadas e usadas no experimento.

Para a classificação do texto o método utilizado foi o Naive Bayes, descrito anteriormente a partir do trabalho de Kang e Yoo (2011). Para medir a performance do sistema utilizou-se métricas padrão de *recall*, precisão e performance e para avaliar a precisão média através das categorias foi utilizado um método de média dado pela Equação (13):

$$F_1(r, p) = \frac{2rp}{r + p} \quad (13)$$

Onde r representa o *recall* e p precisão. A equação visa balancear *recall* e precisão dando-as importâncias iguais.

Para o experimento, foram utilizados um total de 47 categorias e 2.286 documentos coletados na Web. O sistema foi avaliado com as diversas combinações de similaridade calculadas nos passos anteriores e os resultados foram comparados com o método Naive Bayes em um ambiente de aprendizado supervisionado. O sistema descrito por Ko e Seo (2000), obteve um máximo de 71,8% na função de avaliação F_1 e o sistema utilizando o mesmo método em ambiente de aprendizado supervisionado um máximo de 75,6% de performance, desta forma obtendo-se uma diferença de apenas 3,8%. Os autores concluem que a técnica apresentada pode ser um método significativo no aprendizado de texto, uma vez que o método criou automaticamente seu conjunto de treinamento usando uma lista de palavras-chave, e apesar da performance um pouco abaixo do método supervisionado, não se teve o custo de criar manualmente um grande conjunto de treinamento.

2.2.2 Léxico

A maior parte das técnicas de AS podem ser divididas em abordagens de ML e Léxicas. Apesar de que as abordagens de ML tenham obtido avanços significativos na área, aplicá-los requer um conjunto de treinamento rotulado. Conseguir organizar um conjunto de dados de treinamento pode requerer tempo e esforço consideráveis, uma vez que estes dados precisam estar atualizados. A fim de eliminar esta tarefa, são propostos sistemas que geram dados a partir de um conjunto de documentos anotados chamado de corpus (MOREO et al, 2012).

Também chamado de análise baseada em dicionário, esta abordagem, apesar de possuir a vantagem de não fazer uso de um conjunto de treinamento, tem pontos negativos. A maioria dos sistemas por exemplo, utiliza websites com glossários de palavras que podem não conter termos técnicos ou expressões coloquiais. Uma vez que é incapaz de considerar o contexto das expressões, esta técnica não costuma obter uma precisão muito alta para cenários de multi-domínios, onde uma palavra pode expressar uma opinião positiva ou negativa de acordo com o seu contexto (MOREO et al, 2012).

Dentro deste contexto, Kim e Hovy (2004) propõe um sistema de análise de sentimento baseado em dicionário. O sistema tem como objetivo encontrar os sentimentos expressados sobre um dado tópico e identificar a pessoa que o expressou. Para evitar problemas com tons de sentimentos foi especificado que o sistema iria identificar sentimentos positivos, negativos ou neutros. Sentenças que não expressam opinião, apenas estabelecem fatos, foram classificadas em um conjunto separado. Como pode ser observado na Figura 12, o sistema de Kim e Hovy (2014), utiliza a técnica de PLN POS *Tagging* para atribuir as classes de palavras e um website com uma base de sinônimos e antônimos para incrementar o conjunto dicionário de palavras.

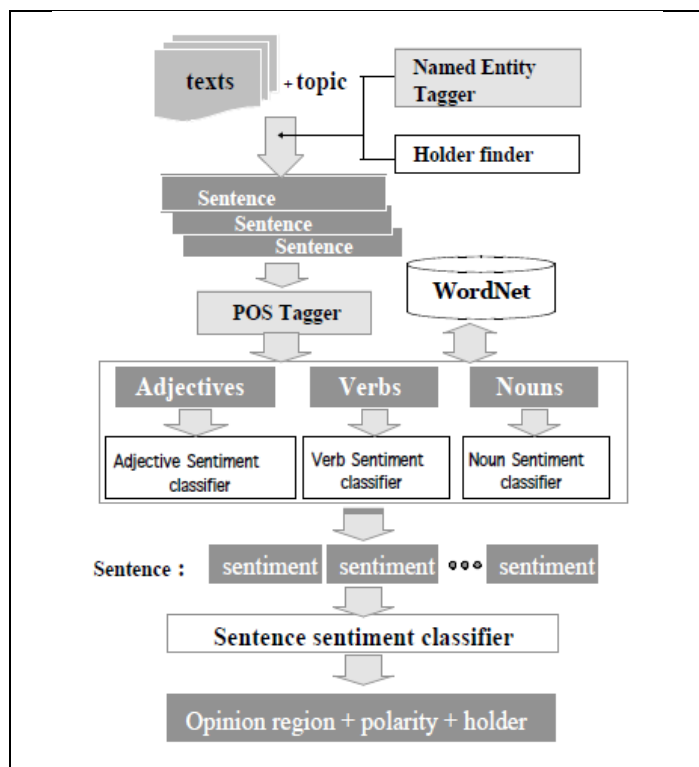


Figura 13. Modelagem do sistema de Kim e Hovy.

Fonte: Kim e Hovy (2004).

Dado um tópico e um conjunto de textos a respeito do tópico, o sistema realiza quatro passos: (i) selecionar as sentenças que contém ambos os tópicos e os detentores de opinião; (ii) delimitar as regiões que se baseiam no detentor da opinião; (iii) calcular a polaridade de todas as palavras que contém sentimento individualmente; e (iv) combinar as seleções e análises anteriores para determinar o sentimento da sentença como um todo (KIM; HOVY, 2004).

Para a classificação de sentimento reuniu-se um pequeno conjunto de palavras manualmente, classificadas por sua polaridade em dois grupos (positivo e negativo), e utilizou-se um website repositório de sinônimos e antônimos para buscar palavras e incrementar os conjuntos. Partindo do princípio que sinônimos de palavras que expressam positividade são positivos e seus antônimos são negativos, antônimos de palavras negativas foram adicionadas ao conjunto das positivas e vice-versa. Esta etapa possibilitou ao dicionário de palavras da aplicação, obter uma quantidade substancial de termos, finalizando um total de 5.880 adjetivos positivos, 6.233 adjetivos negativos, 2.840 verbos positivos e 3.239 verbos negativos.

Tabela 3 - Quantidade de adjetivos e verbos utilizados

	Positivos manuais	Positivos buscados	Negativos manuais	Negativos buscados
Adjetivos	15	5880	19	6233
Verbos	23	2840	21	3239

Fonte: adaptado de Kim e Hovy (2004).

Ao analisar os detentores de opinião, o sistema partiu do princípio de que estes poderiam ser dois: PESSOA e ORGANIZAÇÃO. Para sentenças com mais de um detentor, foi selecionado o detentor mais próximo do tópico. Os modelos de classificação foram divididos em três:

1. Considerando apenas a polaridade do sentimento, sem levar em consideração a sua força.
2. Considera a média harmônica da força do sentimento analisado, que é definida como o número total de classes positivas ou negativas, dividido pela soma do inverso de suas classes. Se uma região contém mais palavras positivas do que negativas, então o sentimento será positivo.
3. Calcula a média geométrica utilizando os mesmos princípios de 1.

Os experimentos foram realizados com a ajuda de três pessoas que classificaram as palavras como sendo positivas ou negativas. O experimento, primeiro comparou os resultados entre as pessoas, a fim de eliminar qualquer disparidade nos testes. E posteriormente comparou os resultados da classificação de palavras do sistema com a classificação humana. Dos dados testados, o sistema classificou 93,07% dos adjetivos e 83,27% dos verbos como sendo positivos ou negativos. O sistema obteve um nível de assertividade de 75,66% com o “humano 1”, 77,88% com o “humano 2”, 81,20% com o “humano 3”.

Os dados foram avaliados com a ajuda de duas pessoas que classificaram 100 sentenças em três categorias (positivo, negativo e “N/A”). As sentenças foram analisadas em quatro grupos distintos: (i) a sentença completa; (ii) palavras entre o detentor da opinião e o tópico; (iii) utilizou o grupo (ii) e mais ou menos duas palavras; e (iv) que utilizou palavras entre o detentor de opinião e o tópico até o fim da sentença. Estes grupos foram analisados separadamente e em conjunto com os modelos listados anteriormente e detectou-se que a melhor média dos modelos foi a do modelo dois, que computa a média harmônica da polaridade das palavras de acordo com a força de seu sentimento.

O sistema obteve uma precisão de 81% com o detentor de opinião fornecido manualmente, e de 67% de precisão com detecção automática do mesmo. Com o experimento, pode-se concluir que apesar das dificuldades e desafios da AS, bons resultados podem ser obtidos, mesmo com modelos relativamente simples e um pequeno conjunto de palavras obtidos manualmente.

Abordagens baseadas em um dicionário de palavras extraem a polaridade de cada sentença em um documento, desta forma o senso de opinião das palavras na frase é processado, a fim de agrupar estas polaridades e classificar o sentimento do texto. Essa área de pesquisa, avalia sentenças e palavras em documentos que expressam opiniões, a fim de extrair sua subjetividade, polaridade e força. Onde a subjetividade é dada caso um texto contenha opiniões ou visões subjetivas. A polaridade visa verificar se um documento expressa uma opinião positiva ou negativa e a força é o grau de polaridade ou de intensidade desta opinião (MOREO et al, 2012).

Moreo et al (2012), propõem um sistema para realizar análise de sentimento de notícias, baseado na técnica léxica de dicionário que efetua as seguintes ações: (i) exclui comentários que contém informações inconclusivas, a fim de deixar a análise mais confiável; (ii) todo o foco implícito ou explícito é automaticamente reconhecido através de um algoritmo (neste trabalho, os autores tratam o tópico ou aspecto do texto como sendo o “foco”, por ser mais apropriado ao contexto em que se refere); (iii) cada foco é avaliado e como resultado se obtém uma classificação de foco determinando a polaridade e a força; e (iv) técnicas de mineração são aplicadas para gerar um sumário interpretável (MOREO et al, 2012).

O sistema proposto visa classificar o documento de maneira geral e também o sentimento de cada tópico de discussão, ou foco. Assim é preciso identificar os tópicos de discussão onde as pessoas expressam sua opinião, para extrair os sentimentos de maneira que não se interfira na análise. Para isto, foi criado manualmente um dicionário estruturado de expressões linguísticas, com base na análise de 250 notícias e este dicionário foi estendido com informação léxica a respeito de assuntos atuais em áreas como política e esportes. Foi aplicado um filtro nas notícias, a fim de excluir todos os comentários inapropriados à análise e ambos o dicionário e os comentários foram pré-processados utilizando técnicas de PLN já abordadas neste trabalho como, *POS Tagging*, derivação de expressões linguísticas e *tokenization*. Seu módulo AS analisa cada sentença dos comentários e cria um conjunto de linhas, representando opiniões de usuários e por fim, estas linhas são utilizadas por um

módulo de mineração de texto, que computa o sentimento geral do texto, bem como o sentimento de cada foco.

Os resultados do sistema foram validados por 20 voluntários, que leram as notícias referentes aos resultados e checaram a consistência das análises. Pode-se verificar que em 87,6% dos casos, os resultados foram completamente consistentes. Em 7,4% dos casos a análise foi insuficiente para determinar com precisão os sentimentos e em 5% os resultados estavam incorretos. O sistema também foi avaliado por seus módulos separadamente e comparado a outros métodos, como os de ML. Obteve-se uma melhor precisão em cenários onde a linguagem coloquial é predominante e apesar de seu conjunto léxico ter sido desenvolvido para a análise de notícias, pode ser adaptado a outros domínios.

2.2.2.1 Uso do PLN em Conjunto com métodos Léxicos

Técnicas de PLN são, por vezes, utilizadas em conjunto com a abordagem léxica, a fim de se pré-processar o texto e encontrar sua estrutura sintática e relações semânticas (MEDHAT; HASSAN; KORASHY, 2014). Pode-se observar este uso das técnicas de PLN em conjunto com as abordagens léxicas de AS nos trabalhos de Kim e Hovy (2014) e no de Moreo et al (2012), onde os autores utilizam das técnicas de *POS Tagging* e *tokenization* nas etapas de pré-processamento de texto, antes de usarem suas técnicas com base na abordagem léxica e proporem seus algoritmos para realizar a AS.

O uso de abordagens léxicas em conjunto com o PLN representa um nível aprofundado de análise, uma vez que estes métodos levam em consideração o contexto ao qual as palavras estão inseridas, incluindo análises por *n-grams*, padrões léxicos, sintáticos e sistemas baseados em regras (DI CARO; GRELLA, 2012). A AS foi inicialmente proposta como uma tarefa de PLN para identificar sentimentos expressos em textos e pode ser utilizada de modo dinâmico com funções de ponderação de cada sentença (TSYT SARAY; PALPANAS, 2011). Desta forma, nota-se que o uso do PLN está diretamente ligado as tarefas de AS, mesmo que de maneira discreta nos métodos de aprendizagem de máquina.

3 DESENVOLVIMENTO

O Capítulo 3 apresenta as especificações do sistema desenvolvido e seu desenvolvimento, as técnicas utilizadas para realizar o PLN e a AS, bem como a análise dos resultados obtidos através da implementação, comparando as opiniões identificadas pelo sistema com a análise humana.

3.1 PROJETO

O objetivo do presente trabalho de conclusão, foi o de desenvolver um sistema que realize a tarefa de AS de um texto inserido na língua portuguesa. Desta forma, propôs-se o desenvolvimento de um sistema para realizar a análise de sentimento através do uso da abordagem Léxica, identificando sentimentos positivos, negativos e neutros, nos textos inseridos.

A escolha do uso da técnica Léxica justificou-se pelo fato de que abordagens *Machine Learning* exigem o uso de um conjunto de treinamento onde as entradas possuem saídas conhecidas (MEDHAT, HASSAN, KORASHY, 2014), o qual não se possuía. Conforme observado por Moreo et al (2012), construir um conjunto de dados de treinamento requer tempo e esforço consideráveis, devido a necessidade destes dados estarem atualizados e a qualidade deste conjunto influenciar diretamente no resultado da análise. As abordagens Léxicas não necessitam da criação destes conjuntos.

Assim propôs-se a utilização das seguintes técnicas de PLN:

1. Segmentação: utilizada para realizar a segmentação das sentenças, separando as frases em suas palavras específicas, visando identificar as palavras que serão utilizadas na tarefa seguinte de classificação;
2. POS *Tagging*: utilizada para atribuir classes específicas as palavras. Identificando as principais classes utilizadas na contagem das palavras positivas e negativas da AS.

A ilustração destas duas etapas pode ser visualizada na Figura 14, logo após a inserção do texto no sistema.

“O Instituto Nacional de Estudos e Pesquisas Anísio Teixeira (INEP) divulgou edital com as normas do Exame Nacional do Ensino Médio ENEM 2017 que possui mudanças em relação à seleção passada. De acordo com o documento, as inscrições serão realizadas entre os dias 8 a 19 de maio, através do site do INEP - <http://enem.inep.gov.br/participante> -. Na inscrição, o candidato deve solicitar atendimento a pessoas deficientes ou sabatistas, além de informar um e-mail para que receba informações referentes ao processo seletivo.

A taxa de participação deste ano passou a ser de R\$82,00, sendo R\$14,00 mais caro do que a edição passada, e que deve ser paga até o dia 24 de maio. A justificativa para o valor da inscrição, de acordo com o Ministério da Educação, o custo por aluno de cada prova foi de R\$91,49 em 2016, ou seja, R\$23,49 a mais do que o valor cobrado na inscrição passada.

Outra mudança é a isenção da taxa de participação, o MEC vai avaliar as informações do candidato com o banco de dados do governo e serão isentos os candidatos que constarem na lista do CadÚnico e os que se enquadram na lei 12.799/2013, aqueles que possuem renda igual ou inferior a um salário mínimo e meio (R\$ 1.405,50). Os candidatos que forem isentos da taxa e faltarem ao exame devem justificar a ausência por meio de documentos, atestados médicos, boletins de ocorrência ou documentos judiciais para ter direito à gratuidade nas outras edições.

As provas serão realizadas nos dias 5 e 12 de novembro, e os locais de aplicação abrirão os portões às 12h até 13h. Meia hora depois, as provas serão entregues aos candidatos. No primeiro dia a prova terá duração de 5h30 para serem respondidas as questões de Linguagens, Códigos e suas Tecnologias, Redação e Ciências da Humanas e suas Tecnologias. No segundo dia, os candidatos terão 4h30 para responderem às questões de Ciências da Natureza e suas Tecnologias e Matemática e suas Tecnologias.

O Enem avalia o desempenho escolar no final do Ensino Médio para selecionar candidatos aptos ao Ensino Superior, através do Sistema de Seleção Unificada (Sisu), que oferece vagas nas Instituições Públicas, e do Programa Universidade para Todos (ProUni), para Instituições Privadas. A nota da prova também pode ser usada para receber benefício do Fundo de Financiamento Estudantil (Fies) para ingresso em cursos técnicos oferecidos pelo Sistema de Seleção Unificada da Educação Profissional e Tecnológica (Sisutec). Nesta edição, a nota do exame não servirá como diploma de conclusão do Ensino Médio.

Qual sua opinião sobre estas mudanças? ”.

Com esta proposta de formulário foi possível coletar os dados. É importante ressaltar, que o sistema busca identificar a polaridade de sentimento expressada de maneira geral por cada opinião encontrada dentro do texto inserido a respeito das mudanças nos padrões de inscrição do ENEM 2017, e não leva em consideração um cenário multi-domínio, onde poderia ser analisado as mudanças em específico, por exemplo, atribuir um sentimento de polaridade para o aumento do valor da inscrição e outro para a mudança na isenção desta taxa. Desta forma, o sistema processa as opiniões através do fluxo ilustrado na Figura 15:

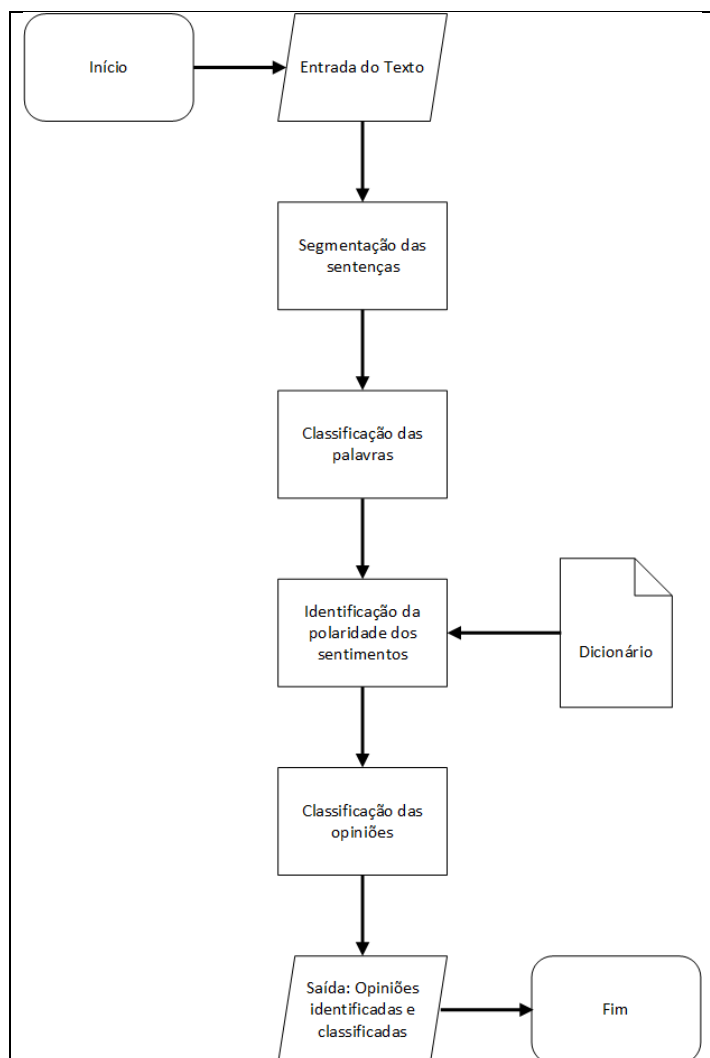


Figura 15. Fluxo de operação do sistema proposto.

Foi possível realizar a classificação das opiniões, que pode ser observada na Figura 15, como o último passo do fluxo antes da saída, através da análise das opiniões identificadas com a implementação do sistema, que identifica várias opiniões dentro de uma mesma frase.

Para efetuar a validação do sistema, a princípio foram excluídas as opiniões coletadas que não estão dentro do contexto proposto, ou seja, não expressam uma opinião passível de análise a respeito do texto das mudanças nas inscrições do ENEM 2017. Posteriormente propôs-se a análise das opiniões coletadas por dois alunos do laboratório LIA, e de uma especialista, que classificaram as opiniões expressadas através dos textos coletados com o formulário como: positivas, negativas ou neutras. O “Aluno 1” identificado neste documento, faz parte do laboratório LIA e é aluno regular do oitavo período do curso de Ciência da Computação da Univali, não possuindo especialização ou curso na área de linguística. O “Aluno 2” identificado neste documento, faz parte do laboratório LIA e é aluno regular do sexto período do curso de Ciência da Computação da Univali, não possuindo especialização ou curso na área de linguística. A “Especialista” identificada neste documento, é formada pelo curso de Publicidade e Propaganda e Especialista em Metodologia da Língua Portuguesa e Literatura pela Uniasselvi - Assevim, possuindo experiência em redação publicitária e literária.

Quanto a qualidade do experimento, esperou-se atingir um nível de precisão médio próximo ao do trabalho similar de Kim e Hovy (2014), sendo este valor um percentual de 79,5%, com uma variação média de 3% entre os alunos avaliadores de opinião do laboratório LIA e a especialista, apresentando assim a análise da qualidade do experimento, que pode ser encontrado na seção 3.3 RESULTADOS.

3.2 DESENVOLVIMENTO

O sistema proposto, foi implementado na linguagem de programação Java. E através da criação de um projeto Apache Maven, foi possível automatizar a compilação e utilizar a biblioteca CoGrOO, que é desenvolvida na mesma linguagem escolhida. O CoGrOO é um Corretor Gramatical acoplável ao LibreOffice, um projeto oficial do CCSL (Centro de Competência em Software Livre) e sua instalação faz com que o LibreOffice passe a detectar erros nas relações entre as palavras, sugerindo uma medida corretiva. Identificando erros gramaticais como colocação pronominal, concordância nominal, concordância entre sujeito e verbo, concordância verbal, entre outros. Com sua adaptação e utilização, foi possível implementar o módulo de PLN.



Figura 16. Logo da biblioteca CoGrOO, utilizada no módulo PLN.

O uso da classe *Analyzer* da biblioteca CoGrOO, permitiu a realização das tarefas de PLN propostas, ou seja, identificar as palavras através da tarefa de segmentação, e realizar a tarefa de POS Tagging, onde é atribuída uma classe gramatical para cada palavra dentro do texto. Conforme a Figura 17, os métodos desta classe, permitiram identificar: (i) os lexemas, ou seja, as palavras que compõem o texto; (ii) seus lemas, que são as palavras em sua forma canônica, por exemplo o lema da palavra “acho” é achar, o da palavra “sendo” é “ser” e assim sucessivamente; e (iii) a classe de cada uma dessas palavras, onde é possível dizer a qual classe gramatical cada palavra pertence, e identificar as classes que possuem importância para a análise, como os adjetivos.

```
Opinião Encontrada.

Palavra: vantajoso.
Lexema: [vantajoso].
Tag: adj.
Carac.: M=S.

Palavra: acho.
Lexema: [achar].
Tag: v-fin.
Carac.: PR=1S=IND.
```

Figura 17. Exemplo de identificação de lemas e classes.

O sistema possui apenas uma tela principal, que pode ser observada na Figura 18, onde é possível selecionar o arquivo com as opiniões a serem analisadas e, ao clicar no botão “Carregar”, o sistema realiza a leitura do arquivo com as opiniões e carrega o dicionário de palavras para comparar com os adjetivos encontrados no texto. O botão “Carregar” também realiza a segmentação das frases lidas a partir do arquivo e a classificação das palavras, bem como a análise da polaridade das opiniões encontradas, e apresenta na parte inferior da tela a saída.

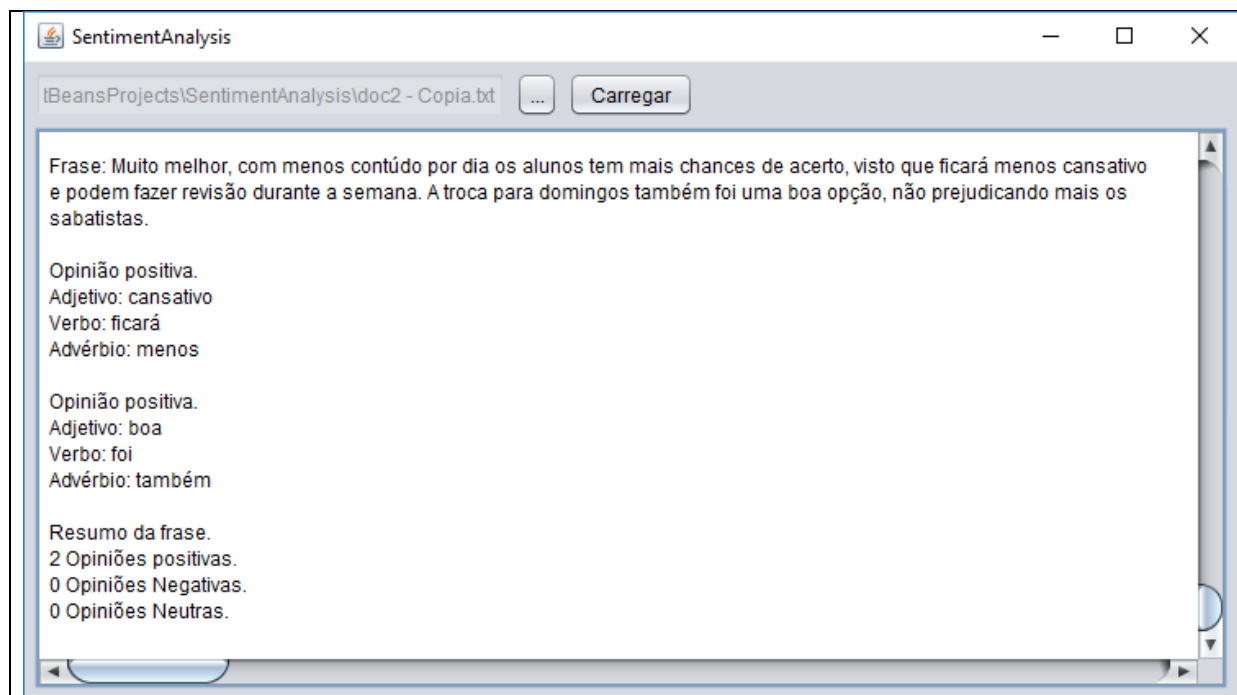


Figura 18. Tela principal do sistema desenvolvido.

Para a criação do dicionário de palavras, foram retirados das opiniões coletadas, dez adjetivos que expressam sentimentos positivos e outros dez que expressam sentimentos negativos, como observado na Tabela 4. Este dicionário de palavras, foi incrementado com a inserção de sinônimos (SINONIMO, 2017) e antônimos (ANTONIMO, 2017) retirados de uma base da Internet. Com este incremento ao dicionário, foi possível estabelecer ao todo 451 adjetivos positivos e 530 adjetivos negativos. A identificação dos sentimentos neutros, foi feita com base nos adjetivos identificados na etapa de *POS Tagging* que não estão reconhecidos como positivos nem como negativos pelo dicionário, desta forma, caso a palavra identificada como um adjetivo não seja polarizada através da comparação com o dicionário, ela é identificada como uma opinião que expressa um sentimento neutro. A leitura das frases coletadas através do formulário, bem como as palavras do dicionário utilizadas para a classificação, foi realizada através de arquivos texto, lidas através da implementação da classe *FileInputStream* do Java.

Tabela 4 - Adjetivos inseridos manualmente no dicionário

	Positivos	Negativos
1	Vantajoso	Cansativo
2	Melhor	Árduo
3	Justo	Injusto
4	Válido	Abandonado
5	Interessante	Absurdo
6	Adequada	Equívoco
7	Favorável	Negativo
8	Coerente	Burocrático
9	Bom	Caro
10	Ideal	Inacessível

Conforme a Figura 19, o sistema desenvolvido, identifica cada opinião expressada dentro de uma determinada frase, e a classifica de acordo com as palavras especificadas no dicionário. Desta forma, sempre que o sistema encontra uma opinião dentro do texto inserido, ele busca a classificação desta opinião dentro do dicionário, comparando o adjetivo encontrado na opinião, com os adjetivos positivos e negativos encontrados no texto. O sistema também considera advérbios de negação, que podem inverter o sentido da opinião expressada, e verbos que caracterizem de fato a opinião, como acho, penso, acredito, ser.

<p>Frase: Dentre todas as características descritas no texto, acho vantajoso as provas serem realizadas em dois domingos consecutivos, pois assim sendo menos cansativo e menos árduo ao candidato que estará prestando o concurso.</p> <p>Sobre o aumento da taxa, pessoalmente não acho correto, pois é um valor elevado para uma prova sendo assim nem sempre acessível a todos.</p> <p>Opinião positiva. Adjetivo: vantajoso Verbo: acho Advérbio: null</p> <p>Opinião positiva. Adjetivo: cansativo Verbo: sendo Advérbio: menos</p> <p>Opinião negativa. Adjetivo: correto Verbo: acho Advérbio: não</p> <p>Opinião negativa. Adjetivo: acessível Verbo: sendo Advérbio: nem</p> <p>Resumo da frase. 2 Opiniões positivas. 2 Opiniões Negativas. 0 Opiniões Neutras.</p>
--

Figura 19. Exemplo de frase com opiniões identificadas.

Para isto o sistema utiliza a seguinte lógica:

1. Percorre o texto inserido buscando um adjetivo, para isto o sistema faz uso da classe gramatical identificada na etapa de PLN com a tarefa de POS *Tagging*;
2. Ao encontrar um adjetivo, o sistema verifica os limites do texto e volta, no máximo quatro posições de palavras, procurando por verbos que caracterizem uma opinião e advérbios que possam inverter o valor das mesmas. Esta etapa, foi testada voltando cinco, quatro e três posições de palavras dentro do texto. Com cinco palavras, a maioria das opiniões atuais que estão sendo analisadas, se confundem com as anteriores, identificando o dobro de opiniões que o texto realmente possui. Com a análise de três posições de palavras, o sistema não identifica a maioria dos advérbios de negação que invertem a polaridade das opiniões como, não, nem, tampouco;
3. Caso o sistema não tenha encontrado um verbo que caracterize a opinião, ele vai adiante do adjetivo encontrado e, checando o limite do texto, percorre no máximo três posições de palavras buscando por um verbo. Esta etapa também foi testada com quatro posições, mas apresentou o mesmo problema que a etapa 2;
4. Se o texto lido contém apenas uma palavra e esta seja um adjetivo, o sistema identifica esta palavra como uma opinião, a fim de compará-la ao dicionário. Caso o sistema encontre um adjetivo que não possua um verbo próximo, não é considerado uma opinião válida;
5. Por fim, o sistema percorre as opiniões, comparando o adjetivo encontrado, com as palavras estabelecidas no dicionário, e se caso o adjetivo esteja indicado no dicionário como um adjetivo positivo e a opinião não possua um advérbio de negação próximo, é considerada uma opinião positiva e vice-versa.

3.3 RESULTADOS

O formulário proposto, possibilitou a coleta de 33 textos, dos quais 30 continham opiniões a respeito das mudanças na inscrição do ENEM 2017. O sistema identificou ao todo, 75 opiniões e a análise das opiniões classificadas pelo sistema, foi feita através da comparação com a classificação realizadas pelos Alunos 1 e 2 do laboratório LIA e da Especialista. Desta forma, foi solicitado que as frases coletadas fossem classificadas da mesma forma que o sistema as classificou, identificando dentro de cada frase as opiniões positivas, negativas e neutras que compunham o texto. Analisando individualmente as classificações realizadas por

cada pessoa, foi possível identificar os acertos na identificação correta das opiniões pelo sistema.

Os melhores resultados foram obtidos com a classificação realizada pela especialista, que identificou, conforme a Tabela 5, 51 opiniões distintas, das quais 46 (90,19%) foram identificadas pelo sistema, com um total de 43 opiniões de acerto geral, representando 84,31% de opiniões corretas identificadas e 93,47% de acerto considerando apenas as opiniões identificadas. O sistema identificou 29 opiniões a mais, porém estas não foram contabilizadas.

Tabela 5 - Resultados obtidos na análise das opiniões pela a especialista

Frase	Especialista				Sistema			
	Positiva	Negativa	Neutra	Total	Positiva	Negativa	Neutra	Acertos
1	2	2	0	4	2	2	0	4
2	1	1	0	2	1	1	0	2
3	2	0	0	2	2	0	0	2
4	1	0	0	1	1	0	0	1
5	1	0	0	1	1	0	0	1
6	0	0	1	1	0	0	1	1
7	1	1	0	2	1	1	0	2
8	2	1	0	3	2	1	0	3
9	1	1	0	2	2	0	0	1
10	1	0	0	1	1	0	0	1
11	1	0	0	1	1	0	0	1
12	0	1	0	1	0	1	0	1
13	4	1	0	5	5	0	0	4
14	1	0	0	1	0	0	0	0
15	0	2	0	2	0	1	0	1
16	1	2	0	3	1	1	1	2
17	1	0	0	1	1	0	0	1
18	1	0	0	1	1	0	0	1
19	0	1	0	1	0	1	0	1
20	0	1	0	1	0	1	0	1
21	1	0	0	1	1	0	0	1
22	0	0	1	1	0	0	1	1
23	1	0	0	1	0	0	0	0
24	0	1	0	1	0	1	0	1
25	2	0	0	2	1	0	0	1
26	0	1	0	1	0	1	0	1
27	1	0	0	1	1	0	0	1
28	2	0	0	2	2	0	0	2
29	3	0	0	3	2	0	0	2
30	2	0	0	2	2	0	0	2
Total	33	16	2	51	31	12	3	43

As opiniões classificadas pelos Alunos 1 e 2, tiveram um percentual de acerto mais ou menos 25% menor do que as identificadas pela especialista. Das 75 opiniões identificadas pelo Aluno 1, 54 opiniões foram identificadas pelo sistema, representando um total de 72%. Destas, 42 estavam corretas, somando um total de 56% de acerto geral e 77,77% de acerto considerando apenas as opiniões identificadas. O sistema identificou 21 opiniões a mais, porém estas não foram contabilizadas. Na Tabela 6 observam-se os resultados da análise do Aluno 1 para cada frase.

Tabela 6 - Resultados obtidos na análise das opiniões identificadas pelo Aluno 1

Frase	Aluno 1				Sistema			
	Positiva	Negativa	Neutra	Total	Positiva	Negativa	Neutra	Acertos
1	2	2	0	4	2	2	0	4
2	1	1	2	4	1	1	1	3
3	2	0	2	4	2	0	0	1
4	0	1	1	2	0	0	2	1
5	1	0	1	2	1	0	0	1
6	0	1	2	3	0	0	1	1
7	1	1	1	3	1	1	0	2
8	2	0	5	7	2	0	3	3
9	1	1	0	2	2	0	0	1
10	1	0	0	1	1	0	0	1
11	1	0	0	1	1	0	0	1
12	1	1	0	2	0	2	0	1
13	2	1	4	7	4	0	1	2
14	0	0	1	1	0	0	0	0
15	0	2	2	4	0	1	1	2
16	1	1	1	3	1	1	1	3
17	1	0	0	1	1	0	0	1
18	1	0	2	3	1	0	0	1
19	1	1	1	3	0	3	0	1
20	0	1	0	1	0	1	0	1
21	1	0	0	1	1	0	0	1
22	0	0	1	1	0	0	1	1
23	0	0	1	1	0	0	0	0
24	0	1	1	2	0	1	0	1
25	0	0	1	1	1	0	0	0
26	0	1	2	3	0	1	0	1
27	1	0	0	1	1	0	0	1
28	2	0	0	2	2	0	0	2
29	3	0	0	3	2	0	0	2
30	2	0	0	2	2	0	0	2
Total	28	16	31	75	29	14	11	42

O Aluno 2 classificou um total de 58 opiniões, das quais 49 foram identificadas pelo sistema, ou seja, 84,48% das opiniões foram identificadas. Dentre as opiniões identificadas

pelo sistema, 34 foram corretas, representando um total de 58,62% de opiniões totais corretas e 69,38% de opiniões corretas dentre as que foram identificadas. Na Tabela 7, apresenta-se o resultado das opiniões identificadas pelo Aluno 2.

Tabela 7 - Resultados obtidos na análise das opiniões identificadas pelo Aluno 2

Frase	Aluno 2				Sistema			
	Positiva	Negativa	Neutra	Total	Positiva	Negativa	Neutra	Acertos
1	2	2	0	4	2	2	0	4
2	1	1	1	3	1	1	1	3
3	2	0	1	3	3	0	0	2
4	0	1	0	1	0	0	1	0
5	0	1	1	2	1	0	0	0
6	1	1	1	3	0	0	1	1
7	1	1	0	2	1	1	0	2
8	1	1	1	3	1	1	1	1
9	1	0	1	2	1	0	1	2
10	1	0	0	1	1	0	0	1
11	0	0	1	1	1	0	0	0
12	0	1	1	2	1	1	0	1
13	1	0	2	3	3	0	0	1
14	1	0	0	1	0	0	0	0
15	0	3	0	3	0	1	0	1
16	0	2	1	3	1	1	1	1
17	1	0	0	1	1	0	0	1
18	0	0	1	1	1	0	0	0
19	0	2	0	2	0	2	0	2
20	0	1	0	1	0	1	0	1
21	1	0	0	1	1	0	0	1
22	0	0	1	1	0	0	1	1
23	0	0	0	0	0	0	0	0
24	0	1	0	1	0	1	0	1
25	1	0	1	2	1	0	0	1
26	0	2	0	2	0	1	0	1
27	1	0	0	1	1	0	0	1
28	2	0	1	3	2	1	0	1
29	1	0	2	3	2	0	0	1
30	2	0	0	2	2	0	0	2
Total	21	20	17	58	28	14	7	34

Ao se comparar o resultado obtido pela análise das frases realizadas pelos Alunos, com a análise da Especialista, observou-se um percentual de acerto menor do que o do sistema. Das 51 opiniões identificadas pela Especialista, o Aluno 1 conseguiu identificar 50 (98,03%) e 40 estavam corretas, representando um percentual de 78,43% de acerto geral e 80% de acerto considerando as opiniões identificadas. Na Tabela 8, pode-se observar a comparação das opiniões identificadas pela Especialista com as identificadas pelo Aluno 1.

Tabela 8 - Resultados obtidos na análise das opiniões classificadas pelo Aluno 1 comparando com as identificadas pela Especialista.

Frase	Especialista				Aluno 1			
	Positiva	Negativa	Neutra	Total	Positiva	Negativa	Neutra	Acertos
1	2	2	0	4	2	2	0	4
2	1	1	0	2	1	1	0	2
3	2	0	0	2	1	0	1	1
4	1	0	0	1	0	0	1	0
5	1	0	0	1	1	0	0	1
6	0	0	1	1	0	0	1	1
7	1	1	0	2	1	1	0	2
8	2	1	0	3	1	0	1	1
9	1	1	0	2	1	1	0	2
10	1	0	0	1	1	0	0	1
11	1	0	0	1	1	0	0	1
12	0	1	0	1	0	1	0	1
13	4	1	0	5	2	1	2	3
14	1	0	0	1	0	0	1	0
15	0	2	0	2	0	2	0	2
16	1	2	0	3	1	1	1	2
17	1	0	0	1	1	0	0	1
18	1	0	0	1	1	0	0	1
19	0	1	0	1	0	1	0	1
20	0	1	0	1	0	1	0	1
21	1	0	0	1	1	0	0	1
22	0	0	1	1	0	0	1	1
23	1	0	0	1	0	0	1	0
24	0	1	0	1	0	1	0	1
25	2	0	0	2	0	0	2	0
26	0	1	0	1	0	1	0	1
27	1	0	0	1	1	0	0	1
28	2	0	0	2	2	0	0	2
29	3	0	0	3	3	0	0	3
30	2	0	0	2	2	0	0	2
Total	33	16	2	51	24	14	12	40

O Aluno 2 conseguiu identificar 46 opiniões das 51, representando um total de 90,19% opiniões identificadas. Destas, 34 estavam corretas, o que representa 66,66% de acerto geral e 73,91% de acerto considerando as opiniões identificadas. Na Tabela 9, é possível verificar a análise dos acertos do Aluno 2, comparando com a análise da Especialista.

Tabela 9 - Resultados obtidos na análise das opiniões classificadas pelo Aluno 2 comparando com as identificadas pela Especialista.

Frase	Especialista				Aluno 2			
	Positiva	Negativa	Neutra	Total	Positiva	Negativa	Neutra	Acertos
1	2	2	0	4	2	2	0	4
2	1	1	0	2	1	1	0	2
3	2	0	0	2	2	0	0	2
4	1	0	0	1	0	0	0	0
5	1	0	0	1	0	0	1	0
6	0	0	1	1	0	0	1	1
7	1	1	0	2	1	1	0	2
8	2	1	0	3	0	1	1	1
9	1	1	0	2	1	0	1	1
10	1	0	0	1	1	0	0	1
11	1	0	0	1	0	0	1	0
12	0	1	0	1	0	1	0	1
13	4	1	0	5	1	0	2	1
14	1	0	0	1	1	0	0	1
15	0	2	0	2	0	2	0	2
16	1	2	0	3	0	2	1	2
17	1	0	0	1	1	0	0	1
18	1	0	0	1	0	0	1	0
19	0	1	0	1	0	1	0	1
20	0	1	0	1	0	1	0	1
21	1	0	0	1	1	0	0	1
22	0	0	1	1	0	0	1	1
23	1	0	0	1	0	0	0	0
24	0	1	0	1	0	1	0	1
25	2	0	0	2	1	0	1	1
26	0	1	0	1	0	1	0	1
27	1	0	0	1	1	0	0	1
28	2	0	0	2	1	0	1	1
29	3	0	0	3	1	0	2	1
30	2	0	0	2	2	0	0	2
Total	33	16	2	51	18	14	14	34

Pôde-se observar que as maiores diferenças se concentraram nas opiniões neutras, identificadas pelos Alunos. A análise da especialista, apontou apenas 2 opiniões neutras e o sistema foi capaz de identifica-las corretamente. O Aluno 1 identificou em sua análise, 31 opiniões neutras, e o sistema foi capaz de acertar apenas 10, já o Aluno 2 identificou 17 opiniões como neutras, e o sistema identificou apenas 6 corretamente. Segundo a Especialista, esta diferença deu-se devido a interpretação dos textos analisados. Em sua análise, a Especialista identificou apenas opiniões que continham conteúdo gramatical que caracterizava opiniões, ou seja, verbos e adjetivos que expressavam a opinião do detentor. Já os Alunos, identificaram opiniões realizando inferências e interpretações com base em conhecimento

comum. Opiniões como “acho justo” que apareceram 8 vezes nos textos, foram identificadas pela Especialista e pelo sistema, como opiniões positivas, uma vez que “justo” é um adjetivo que expressa adequação, sensatez. Os Alunos identificaram estas opiniões como neutras, por interpretarem que elas não têm um significado positivo efetivo.

Dentre as dificuldades encontradas, pode-se destacar frases como a 26 “Assim como tudo neste país, a educação esta falida. (...)”, que pode ser considerada uma opinião negativa, porém devido ao erro gramatical da palavra “está” escrita sem o acento, o sistema identifica um pronome demonstrativo “esta” e a palavra “falida” que deveria ser um adjetivo, torna-se um substantivo feminino, não caracterizando a opinião. Esta frase foi identificada como opinião pela análise dos Alunos, porém não foi identificada pela especialista. A frase 14 “Concordo com as alterações, penso que estão de acordo.”, foi identificada como uma opinião, porém o sistema não foi capaz de identificá-la, uma vez que a palavra “concordo” é classificada pelo sistema como um substantivo, e não um verbo conforme o contexto. O trecho “estão de acordo”, também pode ser interpretado como uma opinião positiva, mas o sistema também não foi capaz de identificar, visto que a classificação da palavra “acordo” é feita também como um substantivo e não como um verbo com lema “concordar”. A análise completa das frases realizada pelo sistema se encontra no Apêndice A deste documento. As análises das frases realizadas pelos Alunos e a Especialista se encontram nos Anexos A, B e C deste documento.

4 CONCLUSÕES

O presente trabalho guiou-se a partir de sua Solução Proposta e seus Objetivos, especificados na Subseção 1.2.1 e Seção 1.3 respectivamente. Identificou as técnicas utilizadas pelo PLN para pré-processar textos, segmenta-los nas respectivas palavras desejadas e atribuir classes as palavras através da técnica de POS *Tagging*, bem como um apanhado geral da análise sintática da língua portuguesa. Descreveu os métodos mais utilizados para realizar a tarefa de AS, abordando as duas principais áreas de pesquisa: o Machine Learning e as abordagens Léxicas.

Com a fundamentação teórica, também foi possível responder a maior parte das perguntas da Subseção 1.2.1, conforme abaixo:

1. Quais são as técnicas mais utilizadas para o processamento de linguagem natural?
Foram identificadas todas as fases de análise do PLN e as técnicas mais utilizadas para esta etapa, como as técnicas de Part-of-Speech Tagging e de segmentação, utilizadas no trabalho.
2. Quais os métodos mais utilizados para se fazer Análise de Sentimento em textos?
Não é possível dizer com certeza quais são os métodos mais utilizados para realizar a AS, porém a análise de trabalhos similares, permitiu identificar diversas técnicas e ajudou a definir a técnica que foi utilizada para o desenvolvimento do trabalho.
3. Que base de dados pode ser utilizada para que se extraia informações relevantes? A análise dos trabalhos similares, também permitiu a escolha por coletar dados e classifica-los, criando um dicionário específico para as informações coletadas.
4. Qual linguagem de programação oferece melhor suporte para implementação? Foi identificada a biblioteca CoGroo, desenvolvida em Java, permitindo a implementação de toda a parte de análise do PLN.
5. Quais sentimentos podem ser extraídos na análise de um documento? Os trabalhos similares serviram de base para se decidir identificar as opiniões positivas, negativas e neutras.

6. Qual a qualidade das análises de trabalhos similares? Através do Capítulo 2, foi possível identificar a qualidade dos trabalhos similares, servindo de base para definir a qualidade que se desejou obter com o presente trabalho.

Através do Capítulo 2, foi possível realizar a fundamentação teórica e comparar trabalhos similares, atendendo aos objetivos específicos:

1. Onde foi realizado a fundamentação teórica a respeito do uso e funcionamento do PLN. Este objetivo foi alcançado através da identificação das técnicas e estratégias de implementação utilizadas no pré-processamento dos textos, onde foi possível definir as técnicas que serão utilizadas para o desenvolvimento do projeto;
2. Realizou-se a fundamentação teórica a respeito da aplicação do PLN para a realização da tarefa de AS. Este objetivo foi alcançado através da comparação de trabalhos similares, que identificou a relação direta do PLN com as técnicas de AS Léxicas;
3. Este objetivo foi alcançado pela análise de trabalhos similares. Através dela foi possível definir a utilização da técnica Léxica de dicionário de palavras para a implementação do sistema proposto, uma vez que a mesma está diretamente ligada ao PLN e não exige o uso de um conjunto de treinamento.

Através do Capítulo 3, que especificou o sistema implementado e apresentou a implementação do mesmo com seus resultados, foi possível atender os objetivos específicos:

4. Foi implementado as especificações realizadas;
5. Analisou-se a implementação realizada, comparando os resultados das opiniões identificadas pelo sistema, com os resultados das análises de dois Alunos do laboratório LIA, bem como a análise dos textos realizada pela a Especialista;

Quanto ao objetivo específico 6, caso se considere a análise da Especialista como sendo correta, pode-se dizer que o trabalho atendeu este objetivo, obtendo um percentual de 84,31% de acerto geral, ficando acima do percentual esperado inicialmente de 79,5%. Com a análise das opiniões da Especialista, comparada a análise dos Alunos, pode-se concluir também que o sistema é capaz de classificar as opiniões melhor que uma pessoa que não

possui conhecimento específico na área de linguística, uma vez que o percentual de acerto geral dos Alunos 1 e 2 foi de 78,43% e 66,66% respectivamente, menor que o do sistema.

Desta forma, a investigação de trabalhos similares teve papel fundamental na elaboração do trabalho, pois a partir dela foi possível identificar os sentimentos que seriam analisados pelo sistema proposto. Assim especificou-se que seriam identificados os sentimentos positivos, negativos e neutros. A importância dos trabalhos similares, deu-se também na escolha do método de implementação, o dicionário Léxico, que foi criado manualmente com um conjunto limitado de palavras e em seguida expandido com o uso de uma base de sinônimos e antônimos. Nesta mesma etapa, também foi possível apontar a qualidade dos trabalhos similares, servindo de apoio para especificar o resultado esperado com a implementação.

O objetivo geral do trabalho foi alcançado, uma vez que o sistema desenvolvido realiza a tarefa de análise de sentimento em textos inseridos na língua portuguesa, e apesar das limitações do trabalho, foi possível identificar e apontar os seguintes trabalhos futuros:

1. Incorporar uma das técnicas de Machine Learning estudadas no Capítulo 2 deste documento, visando a comparação e melhora na qualidade das análises;
2. Utilizar as técnicas de PLN abordadas, bem como as técnicas de AS, em conjunto com as técnicas de identificação de comunidades pesquisadas pelo laboratório LIA;
3. Utilizar uma biblioteca alternativa para realizar as tarefas de PLN, a fim de comparação dos resultados;
4. Implementar o corretor ortográfico da biblioteca CoGroo, antes de realizar as tarefas de POS Tagging e AS, a fim de melhorar o resultado da identificação das opiniões.

REFERÊNCIAS

- ARONOFF, M. **Word formation in generative grammar**. Cambridge. MA: MIT Press, 1976.
- BARKER, Ken. Semi-Automatic Recognition of Semantic Relationships in English Technical Texts. **Department of Computer Science, University of Ottawa**. Canada, 1998.
- BIRD, Steven; KLEIN, Ewan; LOPER, Edward. **Natural Language Processing with Python**. 1.ed. Sebastopol, CA: O'Reilly, 2009.
- CEGALLA, Domingos P. **Novíssima Gramática da Língua Portuguesa**. 48.ed. São Paulo, 2010.
- COVINGTON, Michael A. **Natural Language Processing for Prolog Programmers**. 1.ed. Englewood Cliffs, NJ: Pearson Education, 1994.
- DEVIKA, M. D.; SUNITHA, C.; GANESH, Amal. Sentiment Analysis: A Comparative Study On Different Approaches. **Procedia Computer Science**, India, n. 87, p. 44 49, Maio 2016.
- DI CARO, Luigi; GRELLA, Matteo. Sentiment analysis via dependency parsing. **Computer Standards & Interfaces**, Italy, n. 35, p. 442 453, 2012.
- HSU, Chih-Wei; CHANG, Chih-Chung; LIN, Chih-Jen. A Practical Guide to Support Vector Classification. **Department of Computer Science**. Taiwan, May 2016.
- JURAFSKY, Daniel; MARTIN, James H. **Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition**. New Jersey: Prentice Hall, 2000.
- KANG, Hanhoon; YOO, Seong J.; HAN, Dongil. Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews. **Expert Systems with Applications**, Republic of Korea, n. 39, p. 6000 6010, 2011.
- KIM, Soo-Min; HOVY, Eduard. Determining the Sentiment of Opinions. **Proceedings of the COLING conference**, Geneva, 2004.
- KO, Youngjoong; SEO, Jungyun. **Automatic Text Categorization by Unsupervised Learning**. 2000. Disponível em: <www.aclweb.org/anthology/C00-1066>. Acesso em 04 jun. 2017.
- LAPPIN, S.; LEASS, H. An algorithm for pronominal anaphora resolution. **Computational Linguistics**, n. 20, p. 535 561, 1994.
- LI, Yung-Ming; LI, Tsung-Ying. Deriving market intelligence from microblogs. **Decision Support Systems**. Taiwan, v. 55, p. 206 217, February 2013.
- LIMA, A. **Redação Essencial para Concursos: Descobrindo Segredos**. Rio de Janeiro: Elsevier, 2011

LIN, Hsuan-Tien; LIN, Chih-Jen. A Study on Sigmoid Kernels for SVM and the Training of non-PSD Kernels by SMO-type Methods. **Department of Computer Science and Information Engineering**. Taiwan, March 2003.

LIU, Bing. **Sentiment Analysis and Opinion Mining**. 2012. Disponível em: <<https://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf>>. Acesso em 04 jun. 2017.

LIU, Bing. Sentiment Analysis: A Multi-Faceted Problem. **IEEE Intelligent Systems**, Chicago, v. 25, n. 3, p. 76 80, Aug 2010.

LIU, Bing; HSU, Wynne; MA, Yiming. Integrating Classification and Association Rule Mining. **KDD-98 Proceedings**, Singapore, 1998.

MALETTI, Andreas. Survey: Finite-state technology in natural language processing. **Theoretical Computer Science**, Germany, n. 679, p. 2 16, Maio 2017.

MEDHAT, Walaa; HASSAN, Ahmed; KORASHY, Hoda. Sentiment analysis algorithms and applications: A Survey. **Ain Shams Engineering Journal**, Egito, n. 5, p. 1093 1113, Maio 2014.

MOREO, A.; ROMERO, M.; CASTRO, J.L.; ZURITA, J.M. Lexicon-based Comments-oriented News Sentiment Analyzer system. **Expert Systems with Applications**, Spain, n. 39, p. 9166 9180, 2012.

MOTE, Kevin. Natural Language Processing: A survey. **Computing Research Repository**, USA, n. 499, p. 1 70, Maio 2002.

MÜLLER, Daniel N. **Processamento de Linguagem Natural**. 2003. Disponível em: <<http://www.inf.ufrgs.br/~danielnm/docs/pln.pdf>>. Acesso em: 03 jun. 2017.

PEREIRA, Silvio do Lago. **Processamento de Linguagem Natural**. Disponível em: <www.ime.usp.br/~slago/IA-pln.pdf>. Acesso em: 01 jun. 2017.

ROMASZEWSKI, Michal; GLOMB, Przemyslaw; GAWRON, Piotr. Natural hand gestures for human identification in a Human-Computer Interface. **Institute of Theoretical and Applied Informatics**, Poland, March 2013.

RUS, Teodor. A unified language processing methodology. **Theoretical Computer Science**, USA, n. 281, p. 499 536, Maio 2014.

RUSSEL, Stuart J.; NORVIG, Peter. **Artificial Intelligence: A Modern Approach**. Englewood Cliffs, NJ: Prentice-Hall, 1995.

STANGE, Renata Luiza. **Adaptatividade em Aprendizagem de Máquina: Conceitos e Estudo de Caso**. 2011. Disponível em: <http://www.teses.usp.br/teses/disponiveis/3/3141/tde-02072012-175054/publico/Dissertacao_RLStange_2011_Revisada.pdf>. Acesso em 04 jun. 2017.

SUN, Shiliang.; LUO, Chen; CHEN, Junyu. A review of natural language processing techniques for opinion mining systems. **Information Fusion**, China, n. 36, p. 10 25, Julho 2016.

TSYTSAU, Mikalai; PALPANAS, Themis. Survey on Mining Subjective Data on the Web. **Data Mining and Knowledge Discovery**. USA, v. 24, n. 3, p. 478 514, May 2012.

GLOSSÁRIO

Adjunto	Conjunto, como um anexo. Auxiliar.
Adnominal	Conjunto ao nome. Que modifica um nome.
Afixo	Uma parte que se junta a uma palavra para modificar seu significado. Prefixos e sufixos são afixos.
Ambíguas	Palavras que podem ter mais de um sentido ou significado.
Anáforas	Palavras que substituem um sujeito que apareceu antes na frase, e se referenciam a ele.
Bigrama	Conjunto de duas palavras analisados de uma só vez.
Catáfora	Palavra que substitui um sujeito que ainda não apareceu na frase, se referenciando a ele.
Conjugação	Ato de flexionar um verbo, atribuindo-lhe uma característica de modificação.
Corpus	Conjunto de documentos que servem de base para um estudo.
Disparidade	Falta de semelhança, divergência.
Elipses	Omissões de palavras que não alterem o sentido de uma frase.
Esparso	Espalhado.
Estocástico	Processo probabilístico.
Indução	Consequência retirada dos fatos que estão sendo examinados.
Léxico	Compilação de palavras de uma língua, dicionário.
Morfema	Elemento primitivo, a forma mais simples de uma palavra, irreduzível.
Predicativo	Qualificação de um sujeito.
Prefixo	Aquele que vem antes. Parte da palavra que se põe no princípio para lhe modificar o sentido.
Radical	Da raiz, inseparável. Parte invariável de uma palavra que pode receber afixos.
Recência	Qualidade do que é recente.
Saliência	Que se sobressai. Pode se dizer que os fatores de saliência determinam quais terão a maior prioridade, se sobressairão.
Sintagma	Tratado de algum assunto dividido em classes, coleção.
Sufixo	Sílabas ou letras que se unem ao final de uma palavra para modificar seu significado.

APÊNDICE A. ANÁLISE DAS OPINIÕES REALIZADA PELO SISTEMA

Analisando opiniões.

Frase 1: Dentre todas as características descritas no texto, acho vantajoso as provas serem realizadas em dois domingos consecutivos, pois assim sendo menos cansativo e menos árduo ao candidato que estará prestando o concurso. Sobre o aumento da taxa, pessoalmente não acho correto, pois é um valor elevado para uma prova sendo assim nem sempre acessível a todos.

Opinião positiva.

Adjetivo: vantajoso

Verbo: acho

Advérbio: null

Opinião positiva.

Adjetivo: cansativo

Verbo: sendo

Advérbio: menos

Opinião negativa.

Adjetivo: correto

Verbo: acho

Advérbio: não

Opinião negativa.

Adjetivo: acessível

Verbo: sendo

Advérbio: nem

Resumo da frase.

2 Opiniões positivas.

2 Opiniões Negativas.

0 Opiniões Neutras.

Frase 2: Quanto ao valor de reajuste, o governo deveria ter provisionado e planejado melhor a questão orçamentária. O percentual ficou muito acima da inflação, o que parece injusto. Acho bastante válido a necessidade de comprovação para a isenção da taxa de inscrição e também a divisão das provas em dois dias. Essa divisão alivia a carga e pressão do candidato.

Opinião neutra.

Adjetivo: orçamentária

Verbo: planejado

Advérbio: melhor

Opinião neutra.

Adjetivo: percentual

Verbo: ficou

Advérbio: null

Opinião negativa.

Adjetivo: injusto

Verbo: parece

Advérbio: null

Opinião positiva.

Adjetivo: válido

Verbo: Acho

Advérbio: bastante

Resumo da frase.

1 Opiniões positivas.

1 Opiniões Negativas.

2 Opiniões Neutras.

Frase 3: O aumento do valor da inscrição é justo, já que o participante que não consegue arcar com as despesas pode pedir isenção da taxa. Sobre dividir a prova durante dois domingos, ao invés de um sábado e domingo corridos, é interessante, já que o participante pode se preparar de forma mais adequada para cada tema abordado no dia. Outra característica importante que vale destacar, é que no primeiro dia já serão as questões que antes eram no

segundo dia (as questões de maiores desafios, já que é muita interpretação de texto e exigia muita concentração, e a redação), onde o participante já se encontrava cansado fisicamente e mentalmente.

Opinião positiva.

Adjetivo: justo

Verbo: é

Advérbio: null

Opinião positiva.

Adjetivo: interessante

Verbo: é

Advérbio: null

Opinião positiva.

Adjetivo: importante

Verbo: destacar

Advérbio: null

Opinião positiva.

Adjetivo: primeiro

Verbo: é

Advérbio: null

Opinião neutra.

Adjetivo: segundo

Verbo: eram

Advérbio: antes

Resumo da frase.

4 Opiniões positivas.

0 Opiniões Negativas.

1 Opiniões Neutras.

Frase 4: A alteração burocrática é necessária. Por se tratar de um exame de carácter público, infelizmente somente quando movimentam-se quesitos financeiros que a responsabilidade e/ou interesse são redobrados.

Opinião neutra.

Adjetivo: burocrática

Verbo: é

Advérbio: null

Opinião positiva.

Adjetivo: necessária

Verbo: é

Advérbio: null

Opinião neutra.

Adjetivo: quesitos

Verbo: movimentam

Advérbio: quando

Resumo da frase.

1 Opiniões positivas.

0 Opiniões Negativas.

2 Opiniões Neutras.

Frase 5: Penso que é justo aumentar custos e dividir a prova em dois dias, mas acredito que deveria ser investido mais em segurança..porque aumentando valor ou burocratizando nem aumenta ou diminui a chance de ter fraudes

Opinião positiva.

Adjetivo: justo

Verbo: é

Advérbio: null

Resumo da frase.

1 Opiniões positivas.

0 Opiniões Negativas.

0 Opiniões Neutras.

Frase 6: Penso que até para as inscrições pagas deveria ser avaliado a renda familiar, pois tem muitos que tem condições para pagar mais de uma faculdade, onde acabam ganhando a vaga pelo enem tirando a chance de uma pessoa que não tem condições de pagar

Opinião neutra.

Adjetivo: pagas

Verbo: avaliado

Advérbio: até

Opinião neutra.

Adjetivo: familiar

Verbo: avaliado

Advérbio: null

Resumo da frase.

0 Opiniões positivas.

0 Opiniões Negativas.

2 Opiniões Neutras.

Frase 7: Quanto ao aumento do valor da inscrição, não sou favorável à essa mudança, pois, deve haver algum equívoco por parte do governo quanto ao custo de 92 reais da prova. Em relação a não cobrança da taxa de inscrição para pessoas de baixa renda, eu sou favorável.

Opinião negativa.

Adjetivo: favorável

Verbo: sou

Advérbio: não

Opinião positiva.

Adjetivo: favorável

Verbo: sou

Advérbio: null

Resumo da frase.

1 Opiniões positivas.

1 Opiniões Negativas.

0 Opiniões Neutras.

Frase 8: Em alguns pontos acredito que a alteração é coerente, por exemplo na taxa de inscrição, nada mais justo ter um reajuste, sendo que o reajuste não cobra nem o custo da prova individual. As provas realizadas em dois dias distintos da um folego para o estudante se preparar para estes dias e matérias específicas. Justo aos que forem isentos de pagar a inscrição caso não realizarem a prova, tenham apresentar documentos que compravam sua falta. O ponto negativo fica pelo tempo de prova, os dois dias deveriam ter tempos iguais para realização da prova. Não entendi o porquê do candidato no momento da inscrição deverá solicitar atendimento a pessoas especiais, isto torna um sistema simples em burocrático.

Opinião positiva.

Adjetivo: coerente

Verbo: é

Advérbio: null

Opinião positiva.

Adjetivo: justo

Verbo: ter

Advérbio: mais

Opinião neutra.

Adjetivo: distintos

Verbo: realizadas

Advérbio: null

Opinião neutra.

Adjetivo: isentos

Verbo: forem

Advérbio: null

Opinião negativa.

Adjetivo: negativo

Verbo: fica

Advérbio: null

Opinião neutra.

Adjetivo: iguais

Verbo: ter

Advérbio: null

Opinião positiva.

Adjetivo: especiais

Verbo: solicitar

Advérbio: null

Opinião neutra.

Adjetivo: simples

Verbo: torna

Advérbio: null

Resumo da frase.

3 Opiniões positivas.

1 Opiniões Negativas.

4 Opiniões Neutras.

Frase 9: Acho que o percentual de aumento foi maior do que o devido, mesmo que agora exista a isenção da taxa de participação em alguns casos. Em relação a prova ser separada em dois finais de semanas distintos, acredito que foi uma boa alternativa.

Opinião neutra.

Adjetivo: percentual

Verbo: Acho

Advérbio: null

Opinião positiva.

Adjetivo: maior

Verbo: foi

Advérbio: null

Opinião neutra.

Adjetivo: distintos

Verbo: acredito

Advérbio: null

Opinião positiva.

Adjetivo: boa

Verbo: foi

Advérbio: null

Resumo da frase.

2 Opiniões positivas.

0 Opiniões Negativas.

2 Opiniões Neutras.

Frase 10: Mudanças interessantes. Sou favorável

Opinião positiva.

Adjetivo: favorável

Verbo: Sou

Advérbio: null

Resumo da frase.

1 Opiniões positivas.

0 Opiniões Negativas.

0 Opiniões Neutras.

Frase 11: Acho que as mudanças são válidas no sentido de melhorar a avaliação dos alunos.

Opinião positiva.

Adjetivo: válidas

Verbo: são

Advérbio: null

Resumo da frase.

1 Opiniões positivas.

0 Opiniões Negativas.

0 Opiniões Neutras.

Frase 12: Acho muito caro a taxa de inscrição, deveria ser gratuita ou bancada pela Universidade do aluno. Porém, pelo menos os alunos mais carentes não precisam pagar.

Opinião negativa.

Adjetivo: caro

Verbo: Acho

Advérbio: muito

Opinião positiva.

Adjetivo: gratuita

Verbo: ser

Advérbio: null

Opinião negativa.

Adjetivo: carentes

Verbo: pagar

Advérbio: mais

Resumo da frase.

1 Opiniões positivas.

2 Opiniões Negativas.

0 Opiniões Neutras.

Frase 13: Sobre a alteração no valor de inscrição, acho justo, desde que seja detalhado publicamente os valores que compõem o custo por aluno para aplicação da prova. Sobre a validação dos dados para identificar alunos que realmente devem ser isentados da inscrição, acredito ser o ideal, para evitar fraudes. Também acredito ser justo o fato de que o aluno faltante necessite comprovar com documentos qual o motivo da falta, e caso não seja comprovado, que perca a isenção da taxa nas próximas provas. Sobre dividir a prova em dois finais de semana, não acho que tenha sido uma alteração positiva, pois facilitará fraudes e fará com que os alunos tenham custos maiores, em alguns casos, para a prova. E finalmente sobre a nota da prova não servir como diploma, acredito que seja o correto. A escola é quem deve emitir a certificação de conclusão do ensino médio.

Opinião positiva.

Adjetivo: justo

Verbo: acho

Advérbio: null

Opinião positiva.

Adjetivo: ideal

Verbo: ser

Advérbio: null

Opinião positiva.

Adjetivo: justo

Verbo: ser

Advérbio: Também

Opinião neutra.

Adjetivo: faltante

Verbo: comprovar

Advérbio: null

Opinião positiva.

Adjetivo: positiva

Verbo: sido

Advérbio: null

Opinião positiva.

Adjetivo: maiores

Verbo: tenham

Advérbio: null

Opinião positiva.

Adjetivo: correto

Verbo: seja

Advérbio: null

Resumo da frase.

6 Opiniões positivas.

0 Opiniões Negativas.

1 Opiniões Neutras.

Frase 14: Concordo com as alterações, penso que estão de acordo.

Resumo da frase.

0 Opiniões positivas.

0 Opiniões Negativas.

0 Opiniões Neutras.

Frase 15: Sabemos, e infelizmente não agimos, que o valor apresentado como despesa por aluno não é o valor da despesa certa. Um país movido a suborno e roubos, o mais fraco paga as dívidas. Isenção na inscrição também é inaceitável. Política e educação são caminhos que nunca deveriam se cruzar.

Opinião neutra.

Adjetivo: fraco

Verbo: paga

Advérbio: null

Opinião negativa.

Adjetivo: inaceitável

Verbo: é

Advérbio: também

Resumo da frase.

0 Opiniões positivas.

1 Opiniões Negativas.

1 Opiniões Neutras.

Frase 16: Bom, minha opinião é que o valor é muito caro, mesmo sendo uma grande oportunidade para muitos, mas o valor pra mim deveria ser mais baixo. E também o tempo que os portões ficam abertos, deveria ser maior. O Enem é muito importante, mas algumas regras e principalmente o valor, pra mim, poderiam ser outros!

Opinião negativa.

Adjetivo: caro

Verbo: é

Advérbio: muito

Opinião positiva.

Adjetivo: grande

Verbo: sendo

Advérbio: mesmo

Opinião neutra.

Adjetivo: baixo

Verbo: ser

Advérbio: mais

Opinião positiva.

Adjetivo: maior

Verbo: ser

Advérbio: null

Opinião positiva.

Adjetivo: importante

Verbo: é

Advérbio: muito

Resumo da frase.

3 Opiniões positivas.

1 Opiniões Negativas.

1 Opiniões Neutras.

Frase 17: Acredito que de acordo com as bonificações resultantes as alterações são válidas.

Opinião neutra.

Adjetivo: resultantes

Verbo: são

Advérbio: de_acordo

Opinião positiva.

Adjetivo: válidas

Verbo: são

Advérbio: null

Resumo da frase.

1 Opiniões positivas.

0 Opiniões Negativas.

1 Opiniões Neutras.

Frase 18: Pelo que entendi, essa mudança seria para controlar a isenção de taxas de inscrições para reduzir as ausências nas provas. Certo. Mas, como vão fazer isso? Pra isenção de taxas, deveria se avaliar cada caso, e dependendo do que tivesse ocorrido com o candidato ele não poderia mais pedir taxa de isenção na próxima prova. Seria uma solução. No mais, não achei as mudanças ruins.

Opinião positiva.

Adjetivo: ruins

Verbo: achei

Advérbio: não

Resumo da frase.

1 Opiniões positivas.

0 Opiniões Negativas.

0 Opiniões Neutras.

Frase 19: A taxa para participar da prova é muito cara, porém existe a vantagem da isenção de pagamento para quem não possui renda suficiente. O fato da prova não ser mais aplicada em dias seguidos também é uma vantagem, pois anteriormente era cansativo realizar a prova sábado e domingo.

Opinião negativa.

Adjetivo: cara

Verbo: é

Advérbio: muito

Opinião negativa.

Adjetivo: suficiente

Verbo: possui

Advérbio: não

Opinião negativa.

Adjetivo: cansativo

Verbo: era

Advérbio: anteriormente

Resumo da frase.

0 Opiniões positivas.

3 Opiniões Negativas.

0 Opiniões Neutras.

Frase 20: Acho as alterações completamente absurdas.

Opinião negativa.

Adjetivo: absurdas

Verbo: Acho

Advérbio: completamente

Resumo da frase.

0 Opiniões positivas.

1 Opiniões Negativas.

0 Opiniões Neutras.

Frase 21: Sou favorável as alterações.

Opinião positiva.

Adjetivo: favorável

Verbo: Sou

Advérbio: null

Resumo da frase.

1 Opiniões positivas.

0 Opiniões Negativas.

0 Opiniões Neutras.

Frase 22: Justificáveis

Opinião neutra.

Adjetivo: Justificáveis

Verbo: Achar

Advérbio: null

Resumo da frase.

0 Opiniões positivas.

0 Opiniões Negativas.

1 Opiniões Neutras.

Frase 23: As mudanças favorecem o alunos que poderão fazer a prova com mais tempo e tranquilidade em dois dias.

Resumo da frase.

0 Opiniões positivas.

0 Opiniões Negativas.

0 Opiniões Neutras.

Frase 24: Diante da atual crise política e econômica acho um absurdo o valor cobrado, com este valor daria pra comprar cerveja para 4 encontros ogrinos.

Opinião neutra.

Adjetivo: política

Verbo: acho

Advérbio: null

Opinião positiva.

Adjetivo: econômica

Verbo: acho

Advérbio: null

Opinião negativa.

Adjetivo: absurdo

Verbo: acho

Advérbio: null

Resumo da frase.

1 Opiniões positivas.

1 Opiniões Negativas.

1 Opiniões Neutras.

Frase 25: Entendo ser justo o reajuste do preço da inscrição, levando em conta os dados apresentados no texto, bem como a isenção da taxa para os candidatos que se enquadrem na Lei 12.799/2013. Minha maior crítica é o período de abertura dos portões. Acredito que a abertura deveria acontecer das 11:00 até as 13:00. Em uma cidade pequena, o horário estabelecido pode ser normalmente cumprido, porém em grandes metrópoles, como São Paulo, existem muitas fatores, como o trânsito, que podem resultar em um atraso do participante. No mais, concordo com as alterações expostas no texto.

Opinião positiva.

Adjetivo: justo

Verbo: ser

Advérbio: null

Opinião neutra.

Adjetivo: 12.799/2013

Verbo: enquadrem

Advérbio: null

Opinião positiva.

Adjetivo: maior

Verbo: é

Advérbio: null

Opinião positiva.

Adjetivo: grandes

Verbo: cumprido

Advérbio: porém

Resumo da frase.

3 Opiniões positivas.

0 Opiniões Negativas.

1 Opiniões Neutras.

Frase 26: Assim como tudo neste país, a educação esta falida. De nada adianta uma prova para avaliar a entrada de alguém na universidade se no ensino médio a educação está abandonada? A base foi esquecida e querem valorizar as universidades. De nada adianta um canudo se nao se tem base.

Opinião neutra.

Adjetivo: médio

Verbo: está

Advérbio: null

Opinião negativa.

Adjetivo: abandonada

Verbo: está

Advérbio: null

Resumo da frase.

0 Opiniões positivas.

1 Opiniões Negativas.

1 Opiniões Neutras.

Frase 27: Acho ótimo a mudança nos dias da aplicação da prova

Opinião positiva.

Adjetivo: ótimo

Verbo: Acho

Advérbio: null

Resumo da frase.

1 Opiniões positivas.

0 Opiniões Negativas.

0 Opiniões Neutras.

Frase 28: Acho bom alguns candidatos poderem solicitar isenção da taxa de inscrição, ajuda pessoas carentes. Também acho ideal as provas serem realizadas em dois finais de semana, divide a carga.

Opinião positiva.

Adjetivo: bom

Verbo: Acho

Advérbio: null

Opinião negativa.

Adjetivo: carentes

Verbo: ajuda

Advérbio: null

Opinião positiva.

Adjetivo: ideal

Verbo: acho

Advérbio: Também

Resumo da frase.

2 Opiniões positivas.

1 Opiniões Negativas.

0 Opiniões Neutras.

Frase 29: Concordo que a prova seja feita em dois dias. Quanto aos valores, acho justo que o aumento da inscrição seja proporcional ao aumento da despesa por aluno. A isenção da inscrição para pessoas carentes também é interessante pois torna a a prova mais inclusiva. Vlw flw.

Opinião positiva.

Adjetivo: justo

Verbo: acho

Advérbio: null

Opinião positiva.

Adjetivo: proporcional

Verbo: seja

Advérbio: null

Opinião negativa.

Adjetivo: carentes

Verbo: é

Advérbio: null

Opinião positiva.

Adjetivo: interessante

Verbo: é

Advérbio: também

Resumo da frase.

3 Opiniões positivas.

1 Opiniões Negativas.

0 Opiniões Neutras.

Frase 30: Muito melhor, com menos contúdo por dia os alunos tem mais chances de acerto, visto que ficará menos cansativo e podem fazer revisão durante a semana. A troca para domingos também foi uma boa opção, não prejudicando mais os sabatistas.

Opinião positiva.

Adjetivo: cansativo

Verbo: ficará

Advérbio: menos

Opinião positiva.

Adjetivo: boa

Verbo: foi

Advérbio: também

Resumo da frase.

2 Opiniões positivas.

0 Opiniões Negativas.

0 Opiniões Neutras.

APÊNDICE B. SINÔNIMOS DOS ADJETIVOS POSITIVOS

Sinônimos dos Adjetivos Positivos						
benéfico	prêmio	assim	infalível	compatível	generoso	gostoso
profícuo	Êxito	excepcional	enérgico	congruente	caridoso	apetitoso
proveitoso	Triunfo	atípico	eficiente	consentânea	humano	deleitoso
útil	vencimento	incomum	eficaz	capaz	benévolo	bravo
proficiente	bastante	inesperado	curioso	justa	benevolente	destemido
benfazejo	diverso	inusitado	intrigante	efetiva	indulgente	corajoso
frutuoso	Vário	raro	estranho	adaptada	misericórdioso	intrépido
questuoso	Tanto	extraordinário	esquisito	ajustada	clemente	valente
rentável	numeroso	singular	intriguista	acomodada	piedoso	resoluto
satisfatório	abundante	pertinente	agradável	conformada	caritativo	impávido
rendoso	copioso	valioso	cativante	talhada	benfeitor	meritório
rendável	inúmero	apropriado	encantador	moldada	aprazível	assegurado
pingue	variado	autêntico	simpático	amoldada	afável	garantido
aproveitável	profuso	ajustado	atraente	cômodo	deleitável	firme
prestadio	farto	lídimo	fascinante	positivo	prazeroso	estável
frutífero	múltiplo	correto	divertido	próspero	adaptado	significante
fecundo	excessivo	adequado	envolvente	benigno	educado	considerável
produtivo	exorbitante	lícito	atrativo	apoiador	cortês	salutar
lucrativo	superabundante	legítimo	importante	adepto	delicado	curado
promissor	intenso	legal	relevante	defensor	gentil	sarado
auspicioso	sobejo	veraz	significativo	cultor	honesto	recuperado
conveniente	nímio	vigente	notável	simpatizante	honrado	restabelecido
propício	grande	certo	estimulante	partidário	íntegro	livre
oportuno	imenso	VERDADEIRO	excitante	aderente	digno	perfeito
superior	enorme	verídico	provocativo	ligado	nobre	maravilhoso
supremo	desmedido	acertado	instigante	unido	decente	sublime
maior	colossal	seguro	apropriada	harmonioso	sério	impecável
maioral	gigantesco	potente	oportuna	coeso	virtuoso	imaginário
máximo	demasiado	vigoroso	propícia	conexo	irrepreensível	irreal
primeiro	deveras	saudável	própria	coesivo	disciplinado	fictício

excelente	sobremaneira	forte	conforme	razoável	obediente	fantástico
soberano	sobremodo	robusto	correta	compreensível	competente	quimérico
sumo	assaz	sólido	certa	lógico	expedito	aspiração
superno	sempre	sadio	acertada	racional	apto	sonho
súpero	amiúde	rijo	satisfatória	moderado	hábil	anseio
preferível	nenhum	forçoso	devida	próprio	habilitado	alvo
vantagem	qualquer	são	cabida	harmônico	saboroso	objetivo
vitória	bem	ativo	cabível	bondoso	delicioso	desejo
exemplar	protótipo	arquétipo	padrão	vontade	modelo	

APÊNDICE C. ANTÔNIMOS DOS ADJETIVOS NEGATIVOS – CONSIDERADOS COMO ADJETIVOS POSITIVOS

Antônimos dos Adjetivos Negativos					
ativante	Vulgar	reto	evidente	positivista	decifrável
energético	Claro	apertado	aparente	otimista	aperceptível
vigorante	Aberto	ditosamente	branco	asseverativo	transparente
fortificante	desimpedido	prosperamente	cândido	assertivo	óbvio
tônico	desobstruído	afortunadamente	explícito	vantagoso	tangível
dinâmico	Curto	taxativo	luminoso	apontado	linear
fácil	Abordável	inequívoco	lustroso	assinalado	entendível
motivante	Suave	garantia	luzente	indigitado	perceptível
lúdico	Brando	certeza	luzidio	confiante	intuitivo
leve	Dócil	confiável	conteste	esperançoso	pérvio
tonificante	merecido	exato	comprobatório	afirmativo	dado
convidativo	neutro	preciso	provatório	patriarcal	sociável
acirrate	equitativo	literal	comprovativo	sim	nítido
provocante	reto	rigoroso	comprovante	higiénico	urbano
cordial	direito	lúcido	corroborante	movimentado	oferecido
tentador	devido	patente	assertório	exuberante	conquistável
acessível	puro	perceptível	categórico	esperançado	remediável
confortável	franco	reluzente	confirmador	barato	sanável
leviano	neutral	rútilo	real	econômico	reflexivo
corredio	imparcial	visível	demonstrativo	módico	contagioso
fluente	equável	brilhante	comprovador	parcimonioso	curável
sopa	condigno	notório	edificante	parco	informação
atingível	afinado	níveo	aconselhado	descomplicado	captável
alcançável	retificado	manifesto	indicado	inteligível	possível
consequível	benemérito	cintilante	recomendado	tratável	
lhano	polido	civilizado	pequeno	modesto	
admissível	tolerável	plausível	social	comunicável	
aceitável	expansivo	comunicativo	boa	gratuito	
ótimo	positiva	exíguo	limitado	suportável	

APÊNDICE D. SINÔNIMOS DOS ADJETIVOS NEGATIVOS

Sinônimos dos Adjetivos Negativos						
fatigante	intrincado	lamentavelmente	abate	desprestígia	modera	confusão
estafante	molesto	miseravelmente	desfalca	ameniza	desacerto	atrapalhação
extenuante	intolerável	tristemente	corta	aplaça	equivocação	nunca
exaustivo	insuportável	decrece	gasta	releva	incorrecção	jamais
esgotante	doloroso	abrevia	perde	enfraquece	quiproquó	inverso
fadigoso	acidentado	encurta	dizima	amaina	deslize	desfavorável
trabalhoso	íngreme	reduz	extingue	amortecido	lapso	inconveniente
laborioso	escarpado	resume	sustém	relaxa	falha	contrário
custoso	alcantilado	adelgaça	desgasta	suaviza	erro	nocivo
penoso	fragoso	apouca	morre	acalma	engano	prejudicial
puxado	imerecido	estreita	estraga	abranda	falta	danoso
pesado	indevido	minora	vexa	afrouxa	incerto	contraproducente
aborrecido	desmerecido	encolhe	desmerece	alivia	confuso	maléfico
enfadonhoso	faccioso	abafa	desce	atenua	ambíguo	cético
tedioso	parcial	mingua	dobra	amansa	dúbio	derrotista
chato	iníquo	cai	acanha	esmorece	contestável	pessimista
fastidioso	abusivo	mirra	calca	esmaece	suspeitoso	inútil
desagradável	arbitrário	escasseia	avilta	cede	discutível	nulo
maçante	autoritário	rareia	detrói	cerceia	duvidoso	vão
maçador	despótico	minimiza	desvaloriza	coarcta	suspeito	ineficaz
maçudo	inválido	apequena	deprime	restringe	anfibológico	proibitivo
xarope	desarrazoado	abaixa	humilha	reprime	questionável	suspensório
repetitivo	improcedente	limita	desengrandece	mitiga	controverso	diminutivo
espinhoso	infundado	tira	detrata	reporta	vago	limitativo
duro	injustificável	desconta	deprecia	regula	impreciso	restritivo
difícil	injustificado	deduz	rebaixa	contém	indefinido	suspensivo
complicado	desgraçadamente	subtrai	regateia	adstringe	indefinível	oficial
burocrático	dispendioso	exorbitante	oneroso	salgado	alto	subido
predileto	amado	prezado	querido	estimado	sofrido	precioso
preferido	esquivo	obscuro	insociável	inatingível	impervio	valido

dileto	enigmático	profundo	intratável	inabordá vel	incompreen sível	
--------	------------	----------	------------	-----------------	---------------------	--

APÊNDICE E. ANTÔNIMOS DOS ADJETIVOS POSITIVOS – CONSIDERADOS ADJETIVOS NEGATIVOS

Antônimos dos Adjetivos Positivos						
desvantajoso	levemente	prescrito	detestável	oponente	abstruso	diro
adverso	ligeiramente	irrito	enjoado	contrariador	dissonante	ínfimo
importuno	diminutamente	caduco	indelicado	contestador	assimétrico	mesquinho
desconveniente	raramente	incapaz	ofensivo	oposto	desigual	indecente
hostil	excepcionalmente	abatido	repugnante	infenso	cortado	indigno
mau	invulgarmente	furtivo	descortês	infesto	desagregado	desonesto
péssimo	banalmente	extralegal	obnóxi	opoente	desassociado	desonrado
ruim	comumente	clandestino	inoportuno	disparate	despegado	ímpio
desproveitoso	costumeiramente	anémico	lesivo	tolice	dissociado	sujo
improficuo	habitualmente	quebranto	infrutuoso	kafkiano	inconstante	reles
insatisfatório	normalmente	lânguido	desajustado	despautério	volúvel	desprezível
supérfluo	ilícito	langoroso	inadaptado	despropósito	desunido	insípido
dispensável	enfermo	doentio	inadequado	incôngruo	malvado	insosso
imprestável	ilegal	doente	desadaptado	desconcerto	maligno	dessaboroso
improficiente	impertinente	debilitado	desadequado	disparatado	maldoso	insonso
desnecessário	proibido	alquebrado	impróprio	descabelada	malévolo	insulso
prescindível	débil	entediante	incompatível	inconsistente	malevolente	dessaborido
escusado	quebrado	monótono	incongruente	ridículo	maldito	errado
improdutivo	valetudinário	saturante	despropositado	inconexo	danado	arriscado
infrutífero	morboso	aborrecível	descabido	desconexo	cruel	perigoso
infecundo	morbífero	aborrível	desapropositado	louco	perverso	inseguro
estéril	morbígeno	estopante	desenquadrado	desconchavo	desalmado	malcriado
pior	morbífico	fastiento	inapropriado	desarticulado	impiedoso	deseducado

inferior	mórbido	estopador	incurial	ilógico	despiedoso	grosseiro
último	malsão	enfastioso	incorreto	desorganizado	desapiedado	impolido
nenhum	insalubre	enfastiante	desacertado	patacoada	despiedado	bruto
pouco	enfermiço	enfastiadiço	infausto	aluamento	desumano	incompetente
moderado	vultuoso	desimportante	inimigo	desordenado	frio	inabilitado
parco	mutilado	irrelevante	antagônico	desarmônico	insensível	inapto
mesurado	descalço	desinteressante	adversativo	inconciliável	inumano	inábil
razoável	indisposto	vulgar	resistente	contraditório	terrível	imperito
reduzido	avariado	banal	infeliz	inconsequente	bárbaro	desqualificado
escasso	combalido	trivial	desconforme	desconcordante	bravo	ineficiente
raro	adoentado	ramerraneiro	nefasto	irracional	atroz	imperfeito
diminuto	achacoso	rotineiro	funesto	incoerente	feroz	falho
mínimo	achacado	mediocre	pernicioso	heterogêneo	inclemente	defeituoso
pequeno	improdutível	abominável	nocente	fragmentário	cru	malfeito
insignificante	baldado	repulsivo	daninho	surdo	cruento	desprimoroso
moderadamente	ilegítimo	repelente	sinistro	antilógico	brutal	incompleto
parcamente	franzino	antipático	opositor	variável	celerado	falto
pobremente	entorpecido	rude	contra	separado	implacável	
fracamente	irresoluto	odioso	antagonista	desligado	malfazejo	
minguadamente	frágil	irritante	avesso	irregular	descarido	
cara	inaceitável	abandonada	defectivo	pechoso	menor	

ANEXO A. OPINIÕES ESPECIALISTA

IDENTIFICADAS

PELA

Analisando opiniões.

Frase 1: Dentre todas as características descritas no texto, acho vantajoso as provas serem realizadas em dois domingos consecutivos, pois assim sendo menos cansativo e menos árduo ao candidato que estará prestando o concurso. Sobre o aumento da taxa, pessoalmente não acho correto, pois é um valor elevado para uma prova sendo assim nem sempre acessível a todos.

Opinião positiva.

Adjetivo: vantajoso

Verbo: acho

Advérbio: null

Opinião positiva.

Adjetivo: cansativo

Verbo: sendo

Advérbio: menos

Opinião negativa.

Adjetivo: correto

Verbo: acho

Advérbio: não

Opinião negativa.

Adjetivo: acessível

Verbo: sendo

Advérbio: nem

Resumo da frase.

2 Opiniões positivas.

2 Opiniões Negativas.

0 Opiniões Neutras.

Frase 2: Quanto ao valor de reajuste, o governo deveria ter provisionado e planejado melhor a questão orçamentária. O percentual ficou muito acima da inflação, o que parece injusto. Acho bastante válido a necessidade de comprovação para a isenção da taxa de inscrição e também a divisão das provas em dois dias. Essa divisão alivia a carga e pressão do candidato.

Opinião negativa:

Adjetivo: Injusto

Verbo: parece

Advérbio: muito, acima

Opinião positiva:

Adjetivo: Válido

Verbo: Acho

Advérbio: Bastante

Resumo da frase:

1 positiva

1 negativa

0 neutras

Frase 3: O aumento do valor da inscrição é justo, já que o participante que não consegue arcar com as despesas pode pedir isenção da taxa. Sobre dividir a prova durante dois domingos, ao invés de um sábado e domingo corridos, é interessante, já que o participante pode se preparar de forma mais adequada para cada tema abordado no dia. Outra característica importante que vale destacar, é que no primeiro dia já serão as questões que antes eram no segundo dia (as questões de maiores desafios, já que é muita interpretação de texto e exigia muita concentração, e a redação), onde o participante já se encontrava cansado fisicamente e mentalmente.

Opinião positiva:

Adjetivo: Justo

Verbo: É

Advérbio: X

Opinião positiva:

Adjetivo: Interessante

Verbo: É

Advérbio: Durante

Resumo da frase:

2 positivas

0 negativa

0 neutras

Frase 4: A alteração burocrática é necessária. Por se tratar de um exame de caráter público, infelizmente somente quando movimentam-se quesitos financeiros que a responsabilidade e/ou interesse são redobrados.

Opinião positiva:

Adjetivo: Necessária

Verbo: É

Advérbio: X

Resumo da frase:

1 positiva

0 opiniões negativas

0 neutras

Frase 5: Penso que é justo aumentar custos e dividir a prova em dois dias, mas acredito que deveria ser investido mais em segurança..porque aumentando valor ou burocratizando nem aumenta ou diminui a chance de ter fraudes

Opinião positiva:

Adjetivo: Justo

Verbo: Penso

Advérbio: X

Resumo da frase:

1 positiva

0 negativa

0 neutras

Frase 6: Penso que até para as inscrições pagas deveria ser avaliado a renda familiar, pois tem muitos que tem condições para pagar mais de uma faculdade, onde acabam ganhando a vaga pelo enem tirando a chance de uma pessoa que não tem condições de pagar

Opinião neutra:

Adjetivo: Pagas

Verbo: Penso, deveria

Advérbio: X

Resumo da frase:

0 positiva

0 negativa

1 neutra

Frase 7: Quanto ao aumento do valor da inscrição, não sou favorável à essa mudança, pois, deve haver algum equívoco por parte do governo quanto ao custo de 92 reais da prova. Em relação a não cobrança da taxa de inscrição para pessoas de baixa renda, eu sou favorável.

Opinião negativa:

Adjetivo: Favorável

Verbo: Sou

Advérbio: X

Opinião positiva:

Adjetivo: Favorável

Verbo: Sou

Advérbio: X

Resumo da frase:

1 positiva

1 negativa

0 neutras

Frase 8: Em alguns pontos acredito que a alteração é coerente, por exemplo na taxa de inscrição, nada mais justo ter um reajuste, sendo que o reajuste não cobra nem o custo da prova individual. As provas realizadas em dois dias distintos da um folego para o estudante se preparar para estes dias e matérias específicas. Justo aos que forem isentos de pagar a inscrição caso não realizarem a prova, tenham apresentar documentos que compravam sua falta. O ponto negativo fica pelo tempo de prova, os dois dias deveriam ter tempos iguais para realização da prova. Não entendi o porquê do candidato no momento da inscrição deverá solicitar atendimento a pessoas especiais, isto torna um sistema simples em burocrático.

Opinião positiva:

Adjetivo: Coerente

Verbo: Acredito, é

Advérbio: X

Opinião positiva:

Adjetivo: Justo

Verbo: Ter

Advérbio: Mais

Opinião negativa:

Adjetivo: Negativo

Verbo: Fica

Advérbio: X

Resumo da frase:

2 positiva

1 negativa

0 neutras

Frase 9: Acho que o percentual de aumento foi maior do que o devido, mesmo que agora exista a isenção da taxa de participação em alguns casos. Em relação a prova ser separada em dois finais de semanas distintos, acredito que foi uma boa alternativa.

Opinião negativa:

Adjetivo: maior

Verbo: Acho, foi

Advérbio: X

Opinião positiva:

Adjetivo: Boa

Verbo: Acredito, foi

Advérbio: X

Resumo da frase:

1 positiva

1 negativa

0 neutras

Frase 10: Mudanças interessantes. Sou favorável

Opinião positiva:

Adjetivo: Favorável

Verbo: Sou

Advérbio: X

Resumo da frase:

1 positiva

0 negativa

0 neutras

Frase 11: Acho que as mudanças são válidas no sentido de melhorar a avaliação dos alunos.

Opinião positiva:

Adjetivo: Válidas

Verbo: Acho, são

Advérbio: X

Resumo da frase:

1 positiva

0 negativa

0 neutras

Frase 12: Acho muito caro a taxa de inscrição, deveria ser gratuita ou bancada pela Universidade do aluno. Porém, pelo menos os alunos mais carentes não precisam pagar.

Opinião negativa:

Adjetivo: Caro

Verbo: Acho

Advérbio: X

Resumo da frase:

0 positiva

1 negativa

0 neutras

Frase 13: Sobre a alteração no valor de inscrição, acho justo, desde que seja detalhado publicamente os valores que compõem o custo por aluno para aplicação da prova. Sobre a validação dos dados para identificar alunos que realmente devem ser isentados da inscrição, acredito ser o ideal, para evitar fraudes. Também acredito ser justo o fato de que o aluno faltante necessite comprovar com documentos qual o motivo da falta, e caso não seja comprovado, que perca a isenção da taxa nas próximas provas. Sobre dividir a prova em dois finais de semana, não acho que tenha sido uma alteração positiva, pois facilitará fraudes e fará com que os alunos tenham custos maiores, em alguns casos, para a prova. E finalmente sobre

a nota da prova não servir como diploma, acredito que seja o correto. A escola é quem deve emitir a certificação de conclusão do ensino médio.

Opinião positiva:

Adjetivo: Justo

Verbo: Acho

Advérbio: X

Opinião positiva:

Adjetivo: Ideal

Verbo: Acredito

Advérbio: X

Opinião positiva:

Adjetivo: Justo

Verbo: Acredito

Advérbio: Também

Opinião negativa:

Adjetivo: Justo, positiva

Verbo: Acho

Advérbio: X

Opinião positiva:

Adjetivo: Correto

Verbo: Acredito

Advérbio: X

Resumo da frase:

4 positivas

1 negativa

0 neutras

Frase 14: Concordo com as alterações, penso que estão de acordo.

Opinião positiva:

Adjetivo: X

Verbo: Concordo

Advérbio: X

Resumo da frase:

1 positiva

0 negativa

0 neutras

Frase 15: Sabemos, e infelizmente não agimos, que o valor apresentado como despesa por aluno não é o valor da despesa certa. Um país movido a suborno e roubos, o mais fraco paga as dívidas. Isenção na inscrição também é inaceitável. Política e educação são caminhos que nunca deveriam se cruzar.

Opinião negativa:

Adjetivo: X

Verbo: Sabemos, agimos

Advérbio: Infelizmente

Opinião negativa:

Adjetivo: Inaceitável

Verbo: É

Advérbio: Também

Resumo da frase:

0 positivas

2 negativas

0 neutras

Frase 16: Bom, minha opinião é que o valor é muito caro, mesmo sendo uma grande oportunidade para muitos, mas o valor pra mim deveria ser mais baixo. E também o tempo

que os portões ficam abertos, deveria ser maior. O Enem é muito importante, mas algumas regras e principalmente o valor, pra mim, poderiam ser outros!

Opinião negativa:

Adjetivo: Caro

Verbo: É

Advérbio: Muito

Opinião negativa:

Adjetivo: Baixo

Verbo: Deveria

Advérbio: X

Opinião positiva:

Adjetivo: Importante

Verbo: É

Advérbio: X

Resumo da frase:

1 positiva

2 negativas

0 neutras

Frase 17: Acredito que de acordo com as bonificações resultantes as alterações são válidas.

Opinião positiva:

Adjetivo: Válidas

Verbo: São

Advérbio: X

Resumo da frase:

1 positiva

0 negativas

0 neutras

Frase 18: Pelo que entendi, essa mudança seria para controlar a isenção de taxas de inscrições para reduzir as ausências nas provas. Certo. Mas, como vão fazer isso? Pra isenção de taxas, deveria se avaliar cada caso, e dependendo do que tivesse ocorrido com o candidato ele não poderia mais pedir taxa de isenção na próxima prova. Seria uma solução. No mais, não achei as mudanças ruins.

Opinião positiva:

Adjetivo: Ruins

Verbo: Achei

Advérbio: X

Resumo da frase:

1 positiva

0 negativas

0 neutras

Frase 19: A taxa para participar da prova é muito cara, porém existe a vantagem da isenção de pagamento para quem não possui renda suficiente. O fato da prova não ser mais aplicada em dias seguidos também é uma vantagem, pois anteriormente era cansativo realizar a prova sábado e domingo.

Opinião negativa:

Adjetivo: Cara

Verbo: É

Advérbio: Muito

Resumo da frase:

0 positivas

1 negativa

0 neutras

Frase 20: Acho as alterações completamente absurdas.

Opinião negativa:

Adjetivo: Absurdas

Verbo: Acho

Advérbio: Completamente

Resumo da frase:

0 positivas

1 negativa

0 neutras

Frase 21: Sou favorável as alterações.

Opinião positiva:

Adjetivo: Favorável

Verbo: Sou

Advérbio: X

Resumo da frase:

1 positiva

0 negativas

0 neutras

Frase 22: Justificáveis

Opinião neutra:

Adjetivo: Justificáveis

Verbo: X

Advérbio: X

Resumo da frase:

0 positivas

0 negativa

1 neutra

Frase 23: As mudanças favorecem os alunos que poderão fazer a prova com mais tempo e tranquilidade em dois dias.

Opinião positiva:

Adjetivo: X

Verbo: Favorecem

Advérbio: X

Resumo da frase:

1 positivas

0 negativas

0 neutras

Frase 24: Diante da atual crise política e econômica acho um absurdo o valor cobrado, com este valor daria pra comprar cerveja para 4 encontros ogrinos.

Opinião negativa:

Adjetivo: Absurdo

Verbo: Acho

Advérbio: X

Resumo da frase:

0 positivas

1 negativa

0 neutras

Frase 25: Entendo ser justo o reajuste do preço da inscrição, levando em conta os dados apresentados no texto, bem como a isenção da taxa para os candidatos que se enquadrem na Lei 12.799/2013. Minha maior crítica é o período de abertura dos portões. Acredito que a abertura deveria acontecer das 11:00 até as 13:00. Em uma cidade pequena, o horário estabelecido pode ser normalmente cumprido, porém em grandes metrópoles, como São Paulo, existem muitas fatores, como o trânsito, que podem resultar em um atraso do participante. No mais, concordo com as alterações expostas no texto.

Opinião positiva:

Adjetivo: Justo

Verbo: Entendo, ser

Advérbio: X

Opinião positiva:

Adjetivo: X

Verbo: Concordo

Advérbio: X

Resumo da frase:

2 positivas

0 negativa

0 neutras

Frase 26: Assim como tudo neste país, a educação esta falida. De nada adianta uma prova para avaliar a entrada de alguém na universidade se no ensino médio a educação está abandonada? A base foi esquecida e querem valorizar as universidades. De nada adianta um canudo se nao se tem base.

Opinião negativa.

Adjetivo: Abandonada

Verbo: Adianta, Avaliar, Está

Advérbio: X

Resumo da frase.

0 Opiniões positivas.

1 Opiniões Negativas.

0 Opiniões Neutras.

Frase 27: Acho otimo a mudanca nos dias da aplicacao da prova

Opinião positiva.

Adjetivo: Ótimo

Verbo: Acho

Advérbio: X

Resumo da frase.

1 Opiniões positivas.

0 Opiniões Negativas.

0 Opiniões Neutras.

Frase 28: Acho bom alguns candidatos poderem solicitar isenção da taxa de inscrição, ajuda pessoas carentes. Também acho ideal as provas serem realizadas em dois finais de semana, divide a carga.

Opinião positiva.

Adjetivo: Bom

Verbo: Acho

Advérbio: X

Opinião positiva.

Adjetivo: Ideal

Verbo: Acho

Advérbio: X

Resumo da frase.

2 Opiniões positivas.

0 Opiniões Negativas.

0 Opiniões Neutras.

Frase 29: Concordo que a prova seja feita em dois dias. Quanto aos valores, acho justo que o aumento da inscrição seja proporcional ao aumento da despesa por aluno. A isenção da inscrição para pessoas carentes também é interessante pois torna a a prova mais inclusiva. Vlw flw.

Opinião positiva.

Adjetivo: X

Verbo: Concordo

Advérbio: X

Opinião positiva.

Adjetivo: Justo

Verbo: Acho

Advérbio: X

Opinião positiva.

Adjetivo: Interessante

Verbo: É

Advérbio: X

Resumo da frase.

3 Opiniões positivas.

0 Opiniões Negativas.

0 Opiniões Neutras.

Frase 30: Muito melhor, com menos conteúdo por dia os alunos tem mais chances de acerto, visto que ficará menos cansativo e podem fazer revisão durante a semana. A troca para domingos também foi uma boa opção, não prejudicando mais os sabatistas.

Opinião positiva.

Adjetivo: Melhor

Verbo: X

Advérbio: Muito

Opinião positiva.

Adjetivo: Boa

Verbo: Foi

Advérbio: X

Resumo da frase.

2 Opiniões positivas.

0 Opiniões Negativas.

0 Opiniões Neutras.

ANEXO B. OPINIÕES IDENTIFICADAS PELO ALUNO 1

Analisando opiniões.

Frase 1: Dentre todas as características descritas no texto, acho vantajoso as provas serem realizadas em dois domingos consecutivos, pois assim sendo menos cansativo e menos árduo ao candidato que estará prestando o concurso. Sobre o aumento da taxa, pessoalmente não acho correto, pois é um valor elevado para uma prova sendo assim nem sempre acessível a todos.

Opinião positiva.

Adjetivo: vantajoso

Verbo: acho

Advérbio: null

Opinião positiva.

Adjetivo: cansativo

Verbo: sendo

Advérbio: menos

Opinião negativa.

Adjetivo: correto

Verbo: acho

Advérbio: não

Opinião negativa.

Adjetivo: acessível

Verbo: sendo

Advérbio: nem

Resumo da frase.

2 Opiniões positivas.

2 Opiniões Negativas.

0 Opiniões Neutras.

Frase 2: Quanto ao valor de reajuste, o governo deveria ter provisionado e planejado melhor a questão orçamentária. O percentual ficou muito acima da inflação, o que parece injusto. Acho bastante válido a necessidade de comprovação para a isenção da taxa de inscrição e também a divisão das provas em dois dias. Essa divisão alivia a carga e pressão do candidato.

Opinião neutra.

Adjetivo: provisionado

Verbo: ter

Advérbio:

Opinião positiva.

Adjetivo: válido

Verbo:

Advérbio:

Resumo da frase.

1 Opiniões positivas.

1 Opiniões Negativas.

2 Opiniões Neutras.

Frase 3: O aumento do valor da inscrição é justo, já que o participante que não consegue arcar com as despesas podem pedir isenção da taxa. Sobre dividir a prova durante dois domingos, ao invés de um sábado e domingo corridos, é interessante, já que o participante pode se preparar de forma mais adequada para cada tema abordado no dia. Outra característica importante que vale destacar, é que no primeiro dia já serão as questões que antes eram no segundo dia (as questões de maiores desafios, já que é muita interpretação de texto e exigia muita concentração, e a redação), onde o participante já se encontrava cansado fisicamente e mentalmente.

Opinião positiva.

Adjetivo: Justo

Verbo:

Advérbio:

Opinião positiva.

Adjetivo: adequada

Verbo: preparar

Advérbio:

Resumo da frase.

2 Opiniões positivas.

0 Opiniões Negativas.

2 Opiniões Neutras.

Frase 4: A alteração burocrática é necessária. Por se tratar de um exame de caráter público, infelizmente somente quando movimentam-se quesitos financeiros que a responsabilidade e/ou interesse são redobrados.

Opinião negativa.

Adjetivo: somente

Verbo: movimentam-se/ movimentar

Advérbio: quando

Resumo da frase.

0 Opiniões positivas.

1 Opiniões Negativas.

1 Opiniões Neutras.

Frase 5: Penso que é justo aumentar custos e dividir a prova em dois dias, mas acredito que deveria ser investido mais em segurança..porque aumentando valor ou burocratizando nem aumenta ou diminui a chance de ter fraudes.

Opinião positiva.

Adjetivo: justo

Verbo: aumentar

Advérbio:

Resumo da frase.

1 Opiniões positivas.

0 Opiniões Negativas.

1 Opiniões Neutras.

Frase 6: Penso que até para as inscrições pagas deveria ser avaliado a renda familiar, pois tem muitos que tem condições para pagar mais de uma faculdade, onde acabam ganhando a vaga pelo enem tirando a chance de uma pessoa que não tem condições de pagar

Opinião negativa.

Adjetivo: não

Verbo: tem / ter

Advérbio:

Resumo da frase.

Opiniões positivas.

1 Opiniões Negativas.

2 Opiniões Neutras.

Frase 7: Quanto ao aumento do valor da inscrição, não sou favorável à essa mudança, pois, deve haver algum equívoco por parte do governo quanto ao custo de 92 reais da prova. Em relação a não cobrança da taxa de inscrição para pessoas de baixa renda, eu sou favorável.

Opinião negativa.

Adjetivo: não ; favorável

Verbo: sou / ser

Advérbio:

Opinião positiva.

Adjetivo: favorável

Verbo: sou/ ser

Advérbio:

Resumo da frase.

1 Opiniões positivas.

1 Opiniões Negativas.

1 Opiniões Neutras.

Frase 8: Em alguns pontos acredito que a alteração é coerente, por exemplo na taxa de inscrição, nada mais justo ter um reajuste, sendo que o reajuste não cobra nem o custo da prova individual. As provas realizadas em dois dias distintos dá um fôlego para o estudante se preparar para estes dias e matérias específicas. Justo aos que forem isentos de pagar a inscrição caso não realizarem a prova, tenham apresentar documentos que comprovem sua falta. O ponto negativo fica pelo tempo de prova, os dois dias deveriam ter tempos iguais para realização da prova. Não entendi o porquê do candidato no momento da inscrição deverá solicitar atendimento a pessoas especiais, isto torna um sistema simples em burocrático.

Opinião positiva.

Adjetivo: coerente

Verbo: acredito / acredito

Advérbio:

Opinião positiva.

Adjetivo: justo

Verbo:

Advérbio:

Resumo da frase.

2 Opiniões positivas.

0 Opiniões Negativas.

5 Opiniões Neutras.

Frase 9: Acho que o percentual de aumento foi maior do que o devido, mesmo que agora exista a isenção da taxa de participação em alguns casos. Em relação a prova ser separada em dois finais de semanas distintos, acredito que foi uma boa alternativa.

Opinião negativa.

Adjetivo: devido ; maior

Verbo:

Advérbio:

Opinião positiva.

Adjetivo:

Verbo: acredito / acreditar

Advérbio:

Resumo da frase.

1 Opiniões positivas.

1 Opiniões Negativas.

0 Opiniões Neutras.

Frase 10: Mudanças interessantes. Sou favorável

Opinião positiva.

Adjetivo: favorável

Verbo: sou / ser

Advérbio:

Resumo da frase.

1 Opiniões positivas.

0 Opiniões Negativas.

0 Opiniões Neutras.

Frase 11: Acho que as mudanças são válidas no sentido de melhorar a avaliação dos alunos.

Opinião positiva.

Adjetivo: válidas / válido

Verbo: são / ser

Advérbio:

Resumo da frase.

1 Opiniões positivas.

0 Opiniões Negativas.

0 Opiniões Neutras.

Frase 12: Acho muito caro a taxa de inscrição, deveriam ser gratuita ou bancada pela Universidade do aluno. Porém, pelo menos os alunos mais carentes não precisam pagar.

Opinião negativa.

Adjetivo: caro

Verbo:

Advérbio:

Opinião positiva.

Adjetivo: carentes / carente

Verbo: precisar

Advérbio:

Resumo da frase.

1 Opiniões positivas.

1 Opiniões Negativas.

0 Opiniões Neutras.

Frase 13: Sobre a alteração no valor de inscrição, acho justo, desde que seja detalhado publicamente os valores que compõem o custo por aluno para aplicação da prova. Sobre a validação dos dados para identificar alunos que realmente devem ser isentados da inscrição, acredito ser o ideal, para evitar fraudes. Também acredito ser justo o fato de que o aluno faltante necessite comprovar com documentos qual o motivo da falta, e caso não seja comprovado, que perca a isenção da taxa nas próximas provas. Sobre dividir a prova em dois finais de semana, não acho que tenha sido uma alteração positiva, pois facilitará fraudes e fará com que os alunos tenham custos maiores, em alguns casos, para a prova. E finalmente sobre a nota da prova não servir como diploma, acredito que seja o correto. A escola é quem deve emitir a certificação de conclusão do ensino médio.

Opinião positiva.

Adjetivo: ideal

Verbo: acredito / acreditar

Advérbio:

Opinião negativa.

Adjetivo: não ; positiva / positiva

Verbo: acho / achar

Advérbio:

Opinião positiva.

Adjetivo: correto

Verbo: acredito / acreditar

Advérbio:

Resumo da frase.

2 Opiniões positivas.

1 Opiniões Negativas.

4 Opiniões Neutras.

Frase 14: Concordo com as alterações, penso que estão de acordo.

Resumo da frase.

0 Opiniões positivas.

0 Opiniões Negativas.

1 Opiniões Neutras.

Frase 15: Sabemos, e infelizmente não agimos, que o valor apresentado como despesa por aluno não é o valor da despesa certa. Um país movido a suborno e roubos, o mais fraco paga as dívidas. Isenção na inscrição também é inaceitável. Política e educação são caminhos que nunca deveriam se cruzar.

Opinião negativa.

Adjetivo: não

Verbo: agir

Advérbio: infelizmente

Opinião negativa.

Adjetivo: inaceitável

Verbo: inscrição / inscrever

Advérbio: também

Resumo da frase.

0 Opiniões positivas.

2 Opiniões Negativas.

2 Opiniões Neutras.

Frase 16: Bom, minha opinião é que o valor é muito caro, mesmo sendo uma grande oportunidade para muitos, mas o valor pra mim deveria ser mais baixo. E também o tempo que os portões ficam abertos, deveria ser maior. O Enem é muito importante, mas algumas regras e principalmente o valor, pra mim, poderiam ser outros!

Opinião negativa.

Adjetivo: caro

Verbo:

Advérbio:

Opinião positiva.

Adjetivo: importante

Verbo:

Advérbio:

Resumo da frase.

1 Opiniões positivas.

1 Opiniões Negativas.

1 Opiniões Neutras.

Frase 17: Acredito que de acordo com as bonificações resultantes as alterações são válidas.

Opinião positiva.

Adjetivo: válidas / valido

Verbo: alterações/ alterar

Advérbio:

Resumo da frase.

1 Opiniões positivas.

0 Opiniões Negativas.

0 Opiniões Neutras.

Frase 18: Pelo que entendi, essa mudança seria para controlar a isenção de taxas de inscrições para reduzir as ausências nas provas. Certo. Mas, como vão fazer isso? Pra isenção de taxas, deveria se avaliar cada caso, e dependendo do que tivesse ocorrido com o candidato ele não poderia mais pedir taxa de isenção na próxima prova. Seria uma solução. No mais, não achei as mudanças ruins.

Opinião positiva.

Adjetivo: não ; ruim

Verbo: achar

Advérbio:

Resumo da frase.

1 Opiniões positivas.

0 Opiniões Negativas.

2 Opiniões Neutras.

Frase 19: A taxa para participar da prova é muito cara, porém existe a vantagem da isenção de pagamento para quem não possui renda suficiente. O fato da prova não ser mais aplicada em dias seguidos também é uma vantagem, pois anteriormente era cansativo realizar a prova sábado e domingo.

Opinião negativa.

Adjetivo: cara / caro

Verbo: participar

Advérbio:

Opinião positiva.

Adjetivo: não ; suficiente

Verbo: possui

Advérbio:

Resumo da frase.

1 Opiniões positivas.

1 Opiniões Negativas.

1 Opiniões Neutras.

Frase 20: Acho as alterações completamente absurdas.

Opinião negativa.

Adjetivo: absurdas / absurdo

Verbo: acho / achar

Advérbio: completamente

Resumo da frase.

0 Opiniões positivas.

1 Opiniões Negativas.

0 Opiniões Neutras.

Frase 21: Sou favorável às alterações.

Opinião positiva.

Adjetivo: favorável

Verbo: sou / ser

Advérbio:

Resumo da frase.

1 Opiniões positivas.
 0 Opiniões Negativas.
 0 Opiniões Neutras.

Frase 22: Justificáveis

Resumo da frase.

0 Opiniões positivas.
 0 Opiniões Negativas.
 1 Opiniões Neutras.

Frase 23: As mudanças favorecem o alunos que poderão fazer a prova com mais tempo e tranquilidade em dois dias.

Resumo da frase.

0 Opiniões positivas.
 0 Opiniões Negativas.
 1 Opiniões Neutras.

Frase 24: Diante da atual crise política e econômica acho um absurdo o valor cobrado, com este valor daria pra comprar cerveja para 4 encontros ogrinos.

Opinião negativa.

Adjetivo: absurdo
 Verbo: acho / achar
 Advérbio:

Resumo da frase.

0 Opiniões positivas.
 1 Opiniões Negativas.
 1 Opiniões Neutras.

Frase 25: Entendo ser justo o reajuste do preço da inscrição, levando em conta os dados apresentados no texto, bem como a isenção da taxa para os candidatos que se enquadrem na Lei 12.799/2013. Minha maior crítica é o período de abertura dos portões. Acredito que a abertura deveria acontecer das 11:00 até as 13:00. Em uma cidade pequena, o horário estabelecido pode ser normalmente cumprido, porém em grandes metrópoles, como São Paulo, existem muitos fatores, como o trânsito, que podem resultar em um atraso do participante. No mais, concordo com as alterações expostas no texto.

Resumo da frase.

0 Opiniões positivas.

0 Opiniões Negativas.

1 Opiniões Neutras.

Frase 26: Assim como tudo neste país, a educação esta falida. De nada adianta uma prova para avaliar a entrada de alguém na universidade se no ensino médio a educação está abandonada? A base foi esquecida e querem valorizar as universidades. De nada adianta um canudo se nao se tem base.

Opinião negativa

Adjetivo: abandonada / abandonar

Verbo: educar/ educação

Adverbio:

Resumo da frase.

0 Opiniões positivas.

1 Opiniões Negativas.

2 Opiniões Neutras.

Frase 27: Acho otimo a mudanca nos dias da aplicacao da prova

Opinião: positiva

Adjetivo: otimo

Verbo: acho / achar

Adverbio:

Resumo da frase.

1 Opiniões positivas.

0 Opiniões Negativas.

0 Opiniões Neutras.

Frase 28: Acho bom alguns candidatos poderem solicitar isenção da taxa de inscrição, ajuda pessoas carentes. Também acho ideal as provas serem realizadas em dois finais de semana, divide a carga.

Opinião: positiva

Adjetivo: bom

Verbo : acho

Adverbio :

Opinião: positiva

Adjetivo: ideal

Verbo: acho

Adverbio:

Resumo da frase.

2 Opiniões positivas.

0 Opiniões Negativas.

0 Opiniões Neutras.

Frase 29: Concordo que a prova seja feita em dois dias. Quanto aos valores, acho justo que o aumento da inscrição seja proporcional ao aumento da despesa por aluno. A isenção da inscrição para pessoas carentes também é interessante pois torna a a prova mais inclusiva. Vlw flw.

Opinião: positiva

Adjetivo: concordo

Verbo: seja

Adverbio:

Opinião: positiva

Adjetivo: justo

Verbo: acho

Adverbio:

Opinião: positiva

Adjetivo: interessante

Verbo: tona

Adverbio

Resumo da frase.

2 Opiniões positivas.

0 Opiniões Negativas.

0 Opiniões Neutras.

Frase 30: Muito melhor, com menos contúdo por dia os alunos tem mais chances de acerto, visto que ficará menos cansativo e podem fazer revisão durante a semana. A troca para domingos também foi uma boa opção, não prejudicando mais os sabatistas.

Opinião: positiva

Adjetivo : melhor

Verbo: contudo

Adverbio:

Opinião: positiva

Adjetivo: boa

Verbo: opção

Adverbio:

Resumo da frase.

2 Opiniões positivas.

0 Opiniões Negativas.

0 Opiniões Neutras.

ANEXO C. OPINIÕES IDENTIFICADAS PELO ALUNO 2

Analisando opiniões.

Frase 1: Dentre todas as características descritas no texto, acho vantajoso as provas serem realizadas em dois domingos consecutivos, pois assim sendo menos cansativo e menos árduo ao candidato que estará prestando o concurso. Sobre o aumento da taxa, pessoalmente não acho correto, pois é um valor elevado para uma prova sendo assim nem sempre acessível a todos.

Opinião positiva.

Adjetivo: vantajoso

Verbo: acho

Advérbio: null

Opinião positiva.

Adjetivo: cansativo

Verbo: sendo

Advérbio: menos

Opinião negativa.

Adjetivo: correto

Verbo: acho

Advérbio: não

Opinião negativa.

Adjetivo: acessível

Verbo: sendo

Advérbio: nem

Resumo da frase.

2 Opiniões positivas.

2 Opiniões Negativas.

0 Opiniões Neutras.

Frase 2: Quanto ao valor de reajuste, o governo deveria ter provisionado e planejado melhor a questão orçamentária. O percentual ficou muito acima da inflação, o que parece injusto. Acho bastante válido a necessidade de comprovação para a isenção da taxa de inscrição e também a divisão das provas em dois dias. Essa divisão alivia a carga e pressão do candidato.

Opinião neutra.

Adjetivo: melhor

Verbo: planejado

Advérbio: null

Opinião negativa.

Adjetivo: injusto

Verbo: parece

Advérbio: null

Opinião positiva.

Adjetivo: válido

Verbo: Acho

Advérbio: bastante

Resumo da frase.

1 Opiniões positivas.

1 Opiniões Negativas.

1 Opiniões Neutras.

Frase 3: O aumento do valor da inscrição é justo, já que o participante que não consegue arcar com as despesas pode pedir isenção da taxa. Sobre dividir a prova durante dois domingos, ao invés de um sábado e domingo corridos, é interessante, já que o participante pode se preparar de forma mais adequada para cada tema abordado no dia. Outra característica importante que vale destacar, é que no primeiro dia já serão as questões que antes eram no segundo dia (as questões de maiores desafios, já que é muita interpretação de texto e exigia

muita concentração, e a redação), onde o participante já se encontrava cansado fisicamente e mentalmente.

Opinião positiva.

Adjetivo: justo

Verbo: aumento

Advérbio: null

Opinião positiva.

Adjetivo: interessante

Verbo: dividir

Advérbio: corridos

Opinião neutra.

Adjetivo: importante

Verbo: destacar

Advérbio: null

Resumo da frase.

2 Opiniões positivas.

0 Opiniões Negativas.

1 Opiniões Neutras.

Frase 4: A alteração burocrática é necessária. Por se tratar de um exame de caráter público, infelizmente somente quando movimentam-se quesitos financeiros que a responsabilidade e/ou interesse são redobrados.

Opinião negativa.

Adjetivo: infelizmente

Verbo: movimentam

Advérbio: null

Resumo da frase.

0 Opiniões positivas.

1 Opiniões Negativas.

0 Opiniões Neutras.

Frase 5: Penso que é justo aumentar custos e dividir a prova em dois dias, mas acredito que deveria ser investido mais em segurança..porque aumentando valor ou burocratizando nem aumenta ou diminui a chance de ter fraudes

Opinião neutra.

Adjetivo: justo

Verbo: Penso

Advérbio: null

Opinião negativa.

Adjetivo: investido

Verbo: acredito

Advérbio: null

Resumo da frase.

0 Opiniões positivas.

1 Opiniões Negativas.

1 Opiniões Neutras.

Frase 6: Penso que até para as inscrições pagas deveria ser avaliado a renda familiar, pois tem muitos que tem condições para pagar mais de uma faculdade, onde acabam ganhando a vaga pelo enem tirando a chance de uma pessoa que não tem condições de pagar

Opinião neutra.

Adjetivo: pagas

Verbo: Penso

Advérbio: null

Opinião positiva.

Adjetivo: ganhando

Verbo: acabam

Advérbio: null

Opinião negativa.

Adjetivo: tirando

Verbo: tirando

Advérbio: null

Resumo da frase.

1 Opiniões positivas.

1 Opiniões Negativas.

1 Opiniões Neutras.

Frase 7: Quanto ao aumento do valor da inscrição, sou desfavorável à essa mudança, pois, deve haver algum equívoco por parte do governo quanto ao custo de 92 reais da prova. Em relação a não cobrança da taxa de inscrição para pessoas de baixa renda, eu sou favorável.

Opinião negativa.

Adjetivo: desfavorável

Verbo: sou

Advérbio: null

Opinião positiva.

Adjetivo: favorável

Verbo: sou

Advérbio: null

Resumo da frase.

1 Opiniões positivas.

1 Opiniões Negativas.

0 Opiniões Neutras.

Frase 8: Em alguns pontos acredito que a alteração é coerente, por exemplo na taxa de inscrição, nada mais justo ter um reajuste, sendo que o reajuste não cobra nem o custo da prova individual. As provas realizadas em dois dias distintos dá um folego para o estudante se

preparar para estes dias e matérias específicas. Justo aos que forem isentos de pagar a inscrição caso não realizarem a prova, tenham apresentar documentos que compravam sua falta. O ponto negativo fica pelo tempo de prova, os dois dias deveriam ter tempos iguais para realização da prova. Não entendi o porquê do candidato no momento da inscrição deverá solicitar atendimento a pessoas especiais, isto torna um sistema simples em burocrático.

Opinião neutra.

Adjetivo: coerente

Verbo: acredito

Advérbio: null

Opinião positiva.

Adjetivo: distintos

Verbo: preparar

Advérbio: null

Opinião negativa.

Adjetivo: negativo

Verbo: fica

Advérbio: null

Resumo da frase.

1 Opiniões positivas.

1 Opiniões Negativas.

1 Opiniões Neutras.

Frase 9: Acho que o percentual de aumento foi maior do que o devido, mesmo que agora exista a isenção da taxa de participação em alguns casos. Em relação a prova ser separada em dois finais de semanas distintos, acredito que foi uma boa alternativa.

Opinião neutra.

Adjetivo: maior

Verbo: Acho

Advérbio: null

Opinião positiva.

Adjetivo: boa

Verbo: acredito

Advérbio: null

Resumo da frase.

1 Opiniões positivas.

0 Opiniões Negativas.

1 Opiniões Neutras.

Frase 10: Mudanças interessantes. Sou favorável

Opinião positiva.

Adjetivo: favorável

Verbo: Sou

Advérbio: null

Resumo da frase.

1 Opiniões positivas.

0 Opiniões Negativas.

0 Opiniões Neutras.

Frase 11: Acho que as mudanças são válidas no sentido de melhorar a avaliação dos alunos.

Opinião neutra.

Adjetivo: válidas

Verbo: Acho

Advérbio: null

Resumo da frase.

0 Opiniões positivas.

0 Opiniões Negativas.

1 Opiniões Neutras.

Frase 12: Acho muito caro a taxa de inscrição, deveria ser gratuita ou bancada pela Universidade do aluno. Porém, pelo menos os alunos mais carentes não precisam pagar.

Opinião negativa.

Adjetivo: caro

Verbo: Acho

Advérbio: muito

Opinião neutra.

Adjetivo: gratuita

Verbo: ser

Advérbio: null

Resumo da frase.

0 Opiniões positivas.

1 Opiniões Negativas.

1 Opiniões Neutras.

Frase 13: Sobre a alteração no valor de inscrição, acho justo, desde que seja detalhado publicamente os valores que compõem o custo por aluno para aplicação da prova. Sobre a validação dos dados para identificar alunos que realmente devem ser isentados da inscrição, acredito ser o ideal, para evitar fraudes. Também acredito ser justo o fato de que o aluno faltante necessite comprovar com documentos qual o motivo da falta, e caso não seja comprovado, que perca a isenção da taxa nas próximas provas. Sobre dividir a prova em dois finais de semana, não acho que tenha sido uma alteração positiva, pois facilitará fraudes e fará com que os alunos tenham custos maiores, em alguns casos, para a prova. E finalmente sobre a nota da prova não servir como diploma, acredito que seja o correto. A escola é quem deve emitir a certificação de conclusão do ensino médio.

Opinião neutra.

Adjetivo: justo

Verbo: acho

Advérbio: null

Opinião positiva.

Adjetivo: ideal

Verbo: ser

Advérbio: null

Opinião neutra

Adjetivo: justo

Verbo: ser

Advérbio: null

Resumo da frase.

1 Opiniões positivas.

0 Opiniões Negativas.

2 Opiniões Neutras.

Frase 14: Concordo com as alterações, penso que estão de acordo.

Opinião positiva.

Adjetivo: acordo

Verbo: estão

Adjetivo: null

Resumo da frase.

1 Opiniões positivas.

0 Opiniões Negativas.

0 Opiniões Neutras.

Frase 15: Sabemos, e infelizmente não agimos, que o valor apresentado como despesa por aluno não é o valor da despesa certa. Um país movido a suborno e roubos, o mais fraco paga as dívidas. Isenção na inscrição também é inaceitável. Política e educação são caminhos que nunca deveriam se cruzar.

Opinião negativa.

Adjetivo: null

Verbo: agimos

Advérbio: infelizmente

Opinião negativa.

Adjetivo: inaceitável

Verbo: é

Advérbio: null

Opinião negativa.

Adjetivo: null

Verbo: deveriam

Advérbio: nunca

Resumo da frase.

0 Opiniões positivas.

3 Opiniões Negativas.

0 Opiniões Neutras.

Frase 17: Bom, minha opinião é que o valor é muito caro, mesmo sendo uma grande oportunidade para muitos, mas o valor pra mim deveria ser mais baixo. E também o tempo que os portões ficam abertos, deveria ser maior. O Enem é muito importante, mas algumas regras e principalmente o valor, pra mim, poderiam ser outros!

Opinião negativa.

Adjetivo: caro

Verbo: é

Advérbio: muito

Opinião negativa.

Adjetivo: baixo

Verbo: ser

Advérbio: mais

Opinião neutra.

Adjetivo: importante

Verbo: é

Advérbio: muito

Resumo da frase.

0 Opiniões positivas.

2 Opiniões Negativas.

1 Opiniões Neutras.

Frase 17: Acredito que de acordo com as bonificações resultantes as alterações são válidas.

Opinião positiva.

Adjetivo: válidas

Verbo: são

Advérbio: null

Resumo da frase.

1 Opiniões positivas.

0 Opiniões Negativas.

0 Opiniões Neutras.

Frase 18: Pelo que entendi, essa mudança seria para controlar a isenção de taxas de inscrições para reduzir as ausências nas provas. Certo. Mas, como vão fazer isso? Pra isenção de taxas, deveria se avaliar cada caso, e dependendo do que tivesse ocorrido com o candidato ele não poderia mais pedir taxa de isenção na próxima prova. Seria uma solução. No mais, não achei as mudanças ruins.

Opinião neutra.

Adjetivo: ruins

Verbo: achei

Advérbio: null

Resumo da frase.

0 Opiniões positivas.

0 Opiniões Negativas.

1 Opiniões Neutras.

Frase 19: A taxa para participar da prova é muito cara, porém existe a vantagem da isenção de pagamento para quem não possui renda suficiente. O fato da prova não ser mais aplicada em dias seguidos também é uma vantagem, pois anteriormente era cansativo realizar a prova sábado e domingo.

Opinião negativa.

Adjetivo: cara

Verbo: é

Advérbio: muito

Opinião negativa.

Adjetivo: cansativo

Verbo: era

Advérbio: null

Resumo da frase.

0 Opiniões positivas.

2 Opiniões Negativas.

0 Opiniões Neutras.

Frase 20: Acho as alterações completamente absurdas.

Opinião negativa.

Adjetivo: absurdas

Verbo: Acho

Advérbio: completamente

Resumo da frase.

0 Opiniões positivas.

1 Opiniões Negativas.

0 Opiniões Neutras.

Frase 21: Sou favorável as alterações.

Opinião positiva.

Adjetivo: favorável

Verbo: Sou

Advérbio: null

Resumo da frase.

1 Opiniões positivas.

0 Opiniões Negativas.

0 Opiniões Neutras.

Frase 22: Justificáveis

Opinião neutra.

Adjetivo: justificáveis

Verbo: null

Advérbio: null

Resumo da frase.

0 Opiniões positivas.

0 Opiniões Negativas.

1 Opiniões Neutras.

Frase 23: As mudanças favorecem os alunos que poderão fazer a prova com mais tempo e tranquilidade em dois dias.

Resumo da frase.

0 Opiniões positivas.

0 Opiniões Negativas.

0 Opiniões Neutras.

Frase 24: Diante da atual crise política e econômica acho um absurdo o valor cobrado, com este valor daria pra comprar cerveja para 4 encontros ogrinos.

Opinião negativa.

Adjetivo: absurdo

Verbo: acho

Advérbio: null

Resumo da frase.

0 Opiniões positivas.

1 Opiniões Negativas.

0 Opiniões Neutras.

Frase 25: Entendo ser justo o reajuste do preço da inscrição, levando em conta os dados apresentados no texto, bem como a isenção da taxa para os candidatos que se enquadrem na Lei 12.799/2013. Minha maior crítica é o período de abertura dos portões. Acredito que a abertura deveria acontecer das 11:00 até as 13:00. Em uma cidade pequena, o horário estabelecido pode ser normalmente cumprido, porém em grandes metrópoles, como São Paulo, existem muitas fatores, como o trânsito, que podem resultar em um atraso do participante. No mais, concordo com as alterações expostas no texto.

Opinião positiva.

Adjetivo: justo

Verbo: ser

Advérbio: null

Opinião neutra.

Adjetivo: expostas

Verbo: concordo

Advérbio: null

Resumo da frase.

1 Opiniões positivas.

0 Opiniões Negativas.

1 Opiniões Neutras

Frase 26: Assim como tudo neste país, a educação está falida. De nada adianta uma prova para avaliar a entrada de alguém na universidade se no ensino médio a educação está abandonada? A base foi esquecida e querem valorizar as universidades. De nada adianta um canudo se não se tem base.

Opinião negativa.

Adjetivo: falida

Verbo: está

Advérbio: null

Opinião negativa.

Adjetivo: abandonada

Verbo: está

Advérbio: null

Resumo da frase.

0 Opiniões positivas.

2 Opiniões Negativas.

0 Opiniões Neutras.

Frase 27: Acho ótimo a mudança nos dias da aplicação da prova

Opinião positiva.

Adjetivo: ótimo

Verbo: Acho

Advérbio: null

Resumo da frase.

1 Opiniões positivas.

0 Opiniões Negativas.

0 Opiniões Neutras.

Frase 28: Acho bom alguns candidatos poderem solicitar isenção da taxa de inscrição, ajuda pessoas carentes. Também acho ideal as provas serem realizadas em dois finais de semana, divide a carga.

Opinião positiva.

Adjetivo: bom

Verbo: Acho

Advérbio: null

Opinião positiva.

Adjetivo: carentes

Verbo: ajuda

Advérbio: null

Opinião neutra.

Adjetivo: ideal

Verbo: acho

Advérbio: null

Resumo da frase.

2 Opiniões positivas.

0 Opiniões Negativas.

1 Opiniões Neutras.

Frase 29: Concordo que a prova seja feita em dois dias. Quanto aos valores, acho justo que o aumento da inscrição seja proporcional ao aumento da despesa por aluno. A isenção da inscrição para pessoas carentes também é interessante pois torna a a prova mais inclusiva. Vlw flw.

Opinião neutra.

Adjetivo: feita

Verbo: Concordo

Advérbio: null

Opinião neutra.

Adjetivo: justo

Verbo: acho

Advérbio: null

Opinião positiva.

Adjetivo: interessante

Verbo: é

Advérbio: null

Resumo da frase.

1 Opiniões positivas.

0 Opiniões Negativas.

2 Opiniões Neutras.

Frase 30: É muito melhor, com menos conteúdo por dia os alunos tem mais chances de acerto, visto que ficará menos cansativo e podem fazer revisão durante a semana. A troca para domingos também foi uma boa opção, não prejudicando mais os sabatistas.

Opinião positiva.

Adjetivo: melhor

Verbo: É

Advérbio: Muito

Opinião positiva.

Adjetivo: boa

Verbo: foi

Advérbio: null

Resumo da frase.

2 Opiniões positivas.

0 Opiniões Negativas.

0 Opiniões Neutras.