

UNIVERSIDADE SAGRADO CORAÇÃO

JÉSSICA RODRIGUES DA SILVA

**MINERAÇÃO DE OPINIÕES APLICADA A MÍDIAS
SOCIAIS PARA AS ORGANIZAÇÕES**

BAURU
2015

JÉSSICA RODRIGUES DA SILVA

**MINERAÇÃO DE OPINIÕES APLICADA A MÍDIAS
SOCIAIS PARA AS ORGANIZAÇÕES**

Trabalho de conclusão de curso apresentado ao Centro de Ciências Exatas e Sociais Aplicadas da Universidade do Sagrado Coração como parte dos requisitos para obtenção do título de bacharel em Ciência da Computação, sob orientação do Prof. Me. Patrick Pedreira Silva.

BAURU
2015

S5864m	<p data-bbox="552 1381 893 1417">Silva, Jéssica Rodrigues da</p> <p data-bbox="552 1453 1282 1554">Mineração de opiniões aplicada a mídias sociais para as organizações / Jéssica Rodrigues da Silva. -- 2015. 50f. : il.</p> <p data-bbox="584 1600 1153 1633">Orientadora: Profa. Me. Patrick Pedreira Silva.</p> <p data-bbox="552 1675 1282 1774">Trabalho de Conclusão de Curso (Graduação em Ciência da Computação) – Universidade do Sagrado Coração – Bauru – SP.</p> <p data-bbox="552 1816 1282 1906">1. Mineração de opinião. 2. Extração de informação estratégica. 3. Análise de sentimentos. I. Silva, Patrick Pedreira. II. Título.</p>
--------	--

JÉSSICA RODRIGUES DA SILVA

**MINERAÇÃO DE OPINIÕES APLICADA A MÍDIAS SOCIAIS PARA AS
ORGANIZAÇÕES**

Trabalho de conclusão de curso apresentado ao Centro de Ciências Exatas e Sociais Aplicadas da Universidade do Sagrado Coração como parte dos requisitos para obtenção do título de bacharel em Ciência da Computação, sob orientação do Prof. Me. Patrick Pedreira Silva.

Banca examinadora:

Prof. Me. Patrick Pedreira Silva
Universidade Sagrado Coração

Prof. Dr. Elvio Gilberto da Silva
Universidade Sagrado Coração

Prof. Me. Henrique Pachioni Martins
Universidade Sagrado Coração

Bauru, 27 de novembro de 2015.

AGRADECIMENTOS

Agradeço ao meu orientador Patrick Pedreira, que me auxiliou, apoiou e tornou possível a realização desta pesquisa.

Á Ricardo Henrique Soares, meu esposo que me apoiou, incentivou e me ajudou diretamente com a realização do projeto, muito obrigada.

Á minha família, que esteve ao meu lado, contribuindo e tornando tudo mais fácil.

E agradeço também a todos que direta ou indiretamente colaboraram com conhecimentos, opiniões, sugestões e críticas, contribuindo para a concretização do presente trabalho e para a minha formação.

RESUMO

Diante do cenário atual de grande competitividade entre as empresas, a posse de informações estratégicas tornou-se uma vantagem valiosa para emplacar produtos e marcas. Nos dias atuais, a internet é o meio em que os consumidores mais expressam suas opiniões sobre a experiência com as organizações, expondo sentimentos positivos e negativos, contidos em cada postagem em sites, redes sociais, fóruns, bate papos, blogs e outros. Esse novo comportamento, criou um novo nicho a ser explorado pelas empresas e por profissionais da computação, a fim de extrair informações que permitam saber como a marca está posicionada no mercado, o que ela pode melhorar baseada na opinião de seu consumidor, qual decisão tomar diante do panorama em que está inserida. Para isso, visto a precariedade das ferramentas atualmente disponíveis, faz-se necessário o desenvolvimento de softwares a fim de classificar a polaridade das opiniões sobre marcas, prover informações estratégicas e táticas para reduzir a incerteza associada a uma decisão e expor diretamente a atitude e pensamento dos consumidores. O presente trabalho teve como objetivo a criação de um software que se conecta a rede social Facebook através da conta de uma organização, e classifica as opiniões publicadas em sua fanpage entre positivas, negativas ou neutras. Com a classificação concluída, são apresentados gráficos que mostram informações sobre a atual reputação da marca na rede social mais utilizada do mundo. O software obteve 70% de acertos ao processar um corpus de cem opiniões reais, resultado que revela um bom desempenho e indica que pode ser um aliado das organizações.

Palavras-chave: Mineração de opinião. Extração de informação estratégica. Análise de sentimentos.

ABSTRACT

Given the current scenario of great competition between companies, ownership of strategic information has become a valuable asset to topple products and brands. Nowadays, the Internet is the medium in which more consumers express their views on the experience of organizations, exposing positive and negative feelings, contained in each posting on websites, social networks, forums, chats, blogs and others. This new behavior, created a new niche to be exploited by businesses and computer professionals in order to extract information that let you know how the brand is positioned in the market, which it can improve based on the opinion of its consumer, what decision to take before the scene in which it operates. For this, as the precariousness of the tools currently available, it is necessary to develop software in order to rank the polarity of opinions about brands, providing strategic and tactical information to reduce the uncertainty associated with a decision and directly exposed the attitude and thinking consumer. This paper proposes the creation of software that will connect the social network Facebook through the account of an organization, and categorizes the opinions posted on your fanpage between positive, negative or neutral. With the complete classification, will be presented graphs that show information about the current reputation of the brand in the most used social network in the world. For development, a fictional fanpage and a corpus of opinions previously registered will be used.

Key-words: Opinion mining. Strategic information extraction. Sentiment analysis.

LISTA DE ILUSTRAÇÕES

Figura 1 - Evolução do PLN.....	15
Figura 2 - Arquitetura de um Interpretador de Língua Natural.....	19
Figura 3 – Opinião positiva.....	26
Figura 4 - Opinião Negativa.....	27
Figura 5 - Diagrama de Processos.....	29
Figura 6 - Primeira Fase: Tela de Login.....	30
Figura 7 - Primeira Fase: Listagem de publicações feitas na fanpage.....	30
Figura 8 - Primeira Fase: Visualização/Respostas de Publicações.....	30
Figura 9 - Primeira Fase: Postagem na fanpage através do sistema.....	31
Figura 10 - Segunda Fase: Opiniões classificadas.....	32
Figura 11 - Gráfico de Resultados.....	34
Figura 12 - Precisão do classificador.....	35
Figura 13 - Quantidade de Opiniões por consumidor no ano.....	36
Figura 14 - Polaridade das opiniões no mês.....	37
Figura 15 - Polaridade das opiniões no ano.....	38

SUMÁRIO

1	INTRODUÇÃO	8
2	OBJETIVOS	10
2.1	OBJETIVO GERAL	10
2.2	OBJETIVOS ESPECÍFICOS.....	10
3	INTELIGÊNCIA ARTIFICIAL	11
4	PROCESSAMENTO DE LINGUAGEM NATURAL (PLN)	13
4.1	HISTÓRICO	13
4.2	APLICAÇÕES	16
4.3	SISTEMAS DE PROCESSAMENTO DE LINGUAGEM NATURAL (SPLN)	18
4.3.1	Arquitetura de um Sistema de Interpretação de Linguagem Natural	18
5	MINERAÇÃO DE DADOS	22
5.1	MINERAÇÃO DE OPINIÃO	23
5.1.1	Coleta de Conteúdo	23
5.1.2	Classificação	24
5.1.3	Sumarização dos Resultados	24
6	TRABALHOS CORRELATOS	25
7	METODOLOGIA	26
7.1	CORPUS DE COMENTÁRIOS	26
7.1.1	ARQUITETURA DO SISTEMA	27
7.2	DESENVOLVIMENTO	29
8	RESULTADOS	34
8.1	VISÕES APRESENTADAS PELO SISTEMA	35
9	CONSIDERAÇÕES FINAIS	39
	REFERÊNCIAS.....	40

1 INTRODUÇÃO

Toda a informação produzida e disponibilizada em meios como a internet, são de difícil manejo. Antigamente, poucos tinham a capacidade de produzir informação, enquanto muitos atuavam como meros consumidores. Hoje, muitos constroem e consomem grande quantidade de informação, nos desafiando a desbravar este cenário com os mais variados recursos e metodologias para extrair disso informações relevantes e conclusivas. (LEVY, 2000).

O que uma organização deseja, é saber como sua marca está posicionada no mercado, o que ela pode melhorar baseada na opinião de seu consumidor, qual decisão tomar diante do panorama em que está inserida. (PORTER, 1986). Para isso, visto a precariedade das ferramentas atualmente disponíveis, faz-se necessário o desenvolvimento de softwares a fim de classificar a polaridade das opiniões sobre a marca, prover informações estratégicas e táticas para reduzir a incerteza associada a uma decisão e expor diretamente a atitude e pensamento de seu consumidor. (OLIVEIRA, 2006).

Uma sólida alternativa para analisar milhares de opiniões e comentários sobre uma marca, é a utilização da Análise de Sentimentos (AS) ou Mineração de Opinião. Subárea da Inteligência Artificial, a AS tem por finalidade extrair emoções, opiniões ou pontos de vista em textos de linguagem natural, ou seja, produzidos por nós, em nosso idioma. Logo, a AS aplicada à uma rotina que vasculha uma mídia social em busca de comentários citando a organização, faz com que seja possível obter tais informações tão desejadas pelo mercado capitalista em que as organizações estão inseridas. (DOURADO et al., 2010). Hoje, a maioria das organizações realizam tais atividades com o mínimo de automação ou até manualmente, dependendo muito tempo de seus funcionários e as vezes obtendo informações fragmentadas e inconclusivas.

Considerando então que a produção de informação na internet só tende a crescer, o tema desta pesquisa torna-se importante para o futuro das organizações que desejam manter-se ativas e a frente de seus concorrentes, assim como para o cenário de estudos e pesquisas da Análise de Sentimentos e Mineração de Opinião em sistemas inteligentes.

Nessa pesquisa, foi desenvolvido um software que, através da AS, classifica as opiniões obtidas entre positivas, negativas ou neutras, e gera uma informação

final a respeito da organização. Com essas informações, são gerados gráficos, relatórios e visões, as quais podem auxiliar na tomada de decisão. Tendo ciência de sua reputação na web, a organização torna-se capaz de reverter ou melhorar ainda mais esse cenário, tomando decisões que a favoreça.

2 OBJETIVOS

Apresenta-se nos tópicos abaixo o objetivo geral e os objetivos específicos da pesquisa.

2.1 OBJETIVO GERAL

Desenvolver um software que através da Análise de Sentimentos classifique opiniões e gere informações a respeito da organização.

2.2 OBJETIVOS ESPECÍFICOS

- a) pesquisar sobre a Análise de Sentimentos, suas metodologias, técnicas e limitações;
- b) selecionar uma mídia social para extração de comentários a serem classificados;
- c) construir um Corpus de comentários com opiniões reais a serem analisadas;
- d) desenvolver módulo do software que vasculhará a mídia social escolhida em busca de comentários sobre a organização;
- e) desenvolver rotina que identifique a polaridade dessas opiniões a respeito da organização;
- f) desenvolver gráficos e relatórios para exibição das informações obtidas, possibilitando a tomada de decisão.

3 INTELIGÊNCIA ARTIFICIAL

Inteligência Artificial (IA) é o nome dado a uma das áreas mais complexas e fascinantes da Ciência da Computação. A IA pode ser definida como a capacidade de uma máquina de racionar, agir, decidir, armazenar conhecimento e comunicar-se como um ser humano. (GONGORA, 2007).

Rich (1988) e Sato (2009), em definições mais práticas, definem a IA como sendo o estudo de como fazer com que as máquinas pensem e realizem tarefas em que seres humanos são melhores.

Essas definições chegam a ser simplistas e óbvias diante a complexidade e até mesmo obscuridade da área. Como é possível uma máquina tornar-se inteligente? A partir de qual momento podemos considerá-la inteligente? Qual o alcance desse termo?

Alan Turing, pai da computação e IA, fez o mesmo questionamento: “Podem as máquinas pensar?”. Conseguiu responde-lo criando o teste de Turing, onde uma máquina, para ser considerada de fato inteligente, deveria ser submetida a um teste prático chamado “O Jogo da Imitação”, com quatro participantes: o interrogador fazendo as perguntas, um homem e uma máquina ocultos respondendo-as e um júri avaliando-as. Ao final do teste, o júri escolhe qual das respostas é a do ser humano. Caso a máquina seja escolhida, é considerada inteligente. (TURING, 1950).

No ano de 2014, que marcou 60 anos da morte de Turing, um chatbot imitando um garoto de 13 anos, chamado Eugene Goostman, conseguiu passar no teste convencendo 33% dos juizes de que era humano. A prova aconteceu durante o Turing Test 2014, um evento anual organizado pela Universidade de Reading, na Inglaterra, em que máquinas e softwares tentam passar no teste de Turing. Por ser o primeiro software a ser aprovado, este passa a ser um marco na história da Inteligência Artificial. (ROHR, 2014).

Um dos objetivos da IA é transformar simples computadores em máquinas cognitivas, que na sua forma mais simples, interajam com seres humanos. Se um computador cognitivo pode interagir com o ambiente, logo, poderá atuar sobre ele para melhorá-lo. Atualmente, os seres humanos fornecem respostas incompletas por não terem todas as informações. Já um sistema cognitivo, permitirá a coleta de todas as informações necessárias, incluindo o que é visto, ouvido, sentido e falado, e utilizará esses dados para fornecer respostas mais precisas aos problemas

enfrentados. Uma das linhas da computação cognitiva que será abordada posteriormente é o Processamento de Linguagem Natural (PLN), responsável por tornar possível a comunicação com a máquina através da língua natural do ser humano.

4 PROCESSAMENTO DE LINGUAGEM NATURAL (PLN)

O Processamento de Linguagem Natural (PLN) é a subárea da IA que estuda a capacidade e as limitações de uma máquina em entender a linguagem falada pelos seres humanos no dia a dia. (ROSA, 2011).

O objetivo do Processamento de Linguagem Natural é fornecer aos computadores a capacidade de entender e compor textos. E "entender" um texto significa reconhecer o contexto, fazer análise sintática, semântica, léxica e morfológica, criar resumos, extrair informação, interpretar os sentidos e até aprender conceitos com os textos processados.

No teste de Turing citado anteriormente, um pré-requisito para a máquina participar, é a capacidade de processar línguas naturais a fim de habilitá-la a se comunicar com sucesso na língua humana, no caso o inglês. (RUSSEL; NORVIG, 2004).

O PLN está voltado a três aspectos da comunicação em língua natural:

- a) som: prosódia e fonologia;
- b) estrutura: morfologia e sintaxe;
- c) significado: semântica e pragmática.

A prosódia está relacionada aos padrões de ritmo e entonação da língua. A fonologia está relacionada com o reconhecimento dos sons que compõem as palavras de uma língua. A morfologia estuda a construção das palavras, com seus radicais e afixos, que correspondem a partes estáticas e variantes das palavras, como as inflexões verbais. A sintaxe define a estrutura de uma frase, com base na forma como as palavras se relacionam nessa frase. A semântica associa significado a uma estrutura sintática, em termos dos significados das palavras que a compõem. A pragmática verifica se o significado associado à uma estrutura sintática é realmente o significado mais apropriado no contexto considerado. (ROSA, 2011).

4.1 HISTÓRICO

Graças ao surgimento dos computadores no início dos anos 40, novas frentes de pesquisa nos diversos campos do conhecimento científico tiveram avanços significativos. (SILVA et al., 2007).

Com a necessidade de fazê-los “entenderem” instruções para realizarem tarefas, surgiram as linguagens de programação, que deram vida ao início da comunicação homem-máquina. (SILVA et al., 2007).

No início, as linguagens eram mais próximas da linguagem da máquina e mais distantes da do ser humano. Com o passar do tempo, surgiram linguagens como a Lisp e Prolog, que se destacam por distanciarem-se da representação imposta pela arquitetura do computador. Porém, embora sejam instruções mais inteligíveis que as sequências da linguagem de máquina, ainda não são instruções em linguagem natural. (SILVA et al., 2007).

Com o objetivo de transformar computadores em instrumentos mais acessíveis, a saída foi utilizar interfaces gráficas representacionais. Um objeto gráfico significaria várias linhas de codificação na linguagem da máquina. A prova de que esta alternativa deu certo é que os computadores hoje dispõem de sofisticadas interfaces gráficas, com seus menus, ícones, janelas e cores. Esta estratégia não só resolveu o problema do contato direto com a linguagem da máquina, como também transformou os computadores em máquinas atraentes, fáceis de operar e populares, uma vez que os usuários não precisam mais utilizar comandos avançados e de difícil memorização. (SILVA et al., 2007).

Mesmo com esse avanço no relacionamento homem-máquina, a comunicação via linguagem natural continua sendo um desafio: como criar programas capazes de interpretar mensagens codificadas em linguagem natural e decifrá-las para a linguagem de máquina?

Com o passar dos anos, houve muitas pesquisas e desenvolvimentos nos mais diversos ramos do processamento de linguagem natural, destacando-se a tradução automática, considerada pela maioria como o marco inicial na utilização dos computadores para o estudo das línguas naturais. (SILVA et al., 2007). A evolução do PLN é ilustrada da Figura 1.

Figura 1 - Evolução do PLN.

<p>Década de 50: A Tradução automática</p> <ul style="list-style-type: none"> ▪ sistematização computacional das classes de palavras da gramática tradicional ▪ identificação computacional de poucos tipos de constituintes oracionais <p>Década de 60: Novas aplicações e criação de formalismos</p> <ul style="list-style-type: none"> ▪ primeiros tratamentos computacionais das gramáticas livres de contexto ▪ criação dos primeiros analisadores sintáticos ▪ primeiras formalizações do significado em termos de redes semânticas <p>Década de 70: Consolidação dos estudos do PLN</p> <ul style="list-style-type: none"> ▪ implementação de parcelas das primeiras gramáticas e analisadores sintáticos ▪ busca de formalização de fatores pragmáticos e discursivos <p>Década de 80: Sofisticação dos sistemas</p> <ul style="list-style-type: none"> ▪ desenvolvimento de teorias lingüísticas motivadas pelos estudos do PLN <p>Década de 90: Sistemas baseados em “representações do conhecimento”</p> <ul style="list-style-type: none"> ▪ desenvolvimento de projetos de sistemas de PLN complexos que buscam a integração dos vários tipos de conhecimentos lingüísticos e extralingüísticos e das estratégias de inferência envolvidos nos processos de produção, manipulação e interpretação de objetos lingüísticos
--

Fonte: Silva et al. (2007, p. 8).

Nota: Adaptada pela autora.

De modo geral, no PLN, buscam-se soluções para questões computacionais que requerem o tratamento computacional de uma ou mais línguas naturais, quer sejam escritas ou faladas. Mais precisamente, o PLN dedica-se a propor e desenvolver sistemas computacionais que têm a língua natural escrita como objeto primário. Para tanto, linguistas e cientistas da computação, buscam fundamentos em várias disciplinas: Filosofia da Linguagem, Psicologia, Lógica, Inteligência Artificial, Matemática, Ciência da Computação, Linguística Computacional e Linguística. (SILVA et al., 2007).

Em PLN, os linguistas trabalham em duas frentes: utilizam o computador para desenvolver e validar teorias e dados lingüísticos, e fornecem o conhecimento necessário para o desenvolvimento de sistemas especializados. Os cientistas da computação, por sua vez, implementam ferramentas para desenvolvimento e validação dessas teorias, originando os Sistemas de Processamento de Linguagem Natural, mais detalhadamente abordados no tópico 4.3. (SILVA et al., 2007).

4.2 APLICAÇÕES

O atual estágio de desenvolvimento do PLN já permite sua aplicação em áreas como: Extração de Informação, Recuperação de Informação, Tradução Automática, Geração Automática de Texto, Geração de Linguagem Natural, Simplificação de Texto, Correção ortográfica e a que diz respeito a esta pesquisa, Interpretação de Linguagem Natural. Abaixo segue uma breve explanação sobre cada uma:

- a) Extração de Informação (EI): Processo no qual informações são apresentadas pela relevância da busca. A EI parte do princípio de que algumas páginas da Web que tratam de assuntos mais específicos tendem a apresentar regularidade quanto à formatação, estrutura e conteúdo podendo ser agrupadas formando classes de páginas, por exemplo, páginas de cinema, classificados ou eventos científicos. A EI extrai informações relevantes podendo tanto classificar uma página segundo um contexto de domínio como também extrair informações relevantes a este contexto estruturando as informações contidas na página e armazenando-as em bases de dados. (SILVA, 2003);
- b) Recuperação de Informação (RI): RI é a área da Ciência da Computação que permite o acesso fácil e rápido a informações. A RI trata da representação, armazenamento, organização e acesso a tens de informação, de forma organizada e eficiente. (SILVA, 2003);
- c) Tradução Automática (TA): Consiste no processo de transposição de palavras entre idiomas naturais, através da utilização de programas de leitura e interpretação de textos. A Tradução Automática (TA) é um dos domínios da Linguística computacional (LC) que mais envolve conhecimento linguístico, por codificar informações de uma língua para outra. (SILVA et al., 2007);
- d) Geração Automática de Texto: Consiste no processamento de informações já existentes, mas dessa vez de modo mais estruturado e com possibilidade de adequação/manipulação para melhor armazenamento/utilização. (PARDO, 2008);
- e) Geração de Linguagem Natural: Área ainda iniciante na Ciência da Computação, procura construir sistemas programados de linguagem

que se aproxime da Linguagem Natural humana. Tem como foco a interpretação de perguntas feitas em linguagem natural dentro de sistemas de apoio à decisão, reconhecendo estruturas semântica se transformando-as em consultas que retornam resultados relativos à questão elaborada pelo analista. (SILVA et al., 2007);

- f) Interpretação de Linguagem Natural: Consiste no processo através do qual o ser humano interage linguisticamente com a máquina, a qual é provida de programação que lhe permite interpretar enunciados dúbios sem ruídos que possam prejudicar a comunicação. O processamento de Linguagem Natural vem a facilitar a interação do software (através de sua interface) com o usuário, para que se torne mais fácil a comunicação e a passagem de conhecimento, assim quem fizer o uso de um software, possa compreender o que ele tem a oferecer e consiga saber o que o usuário está necessitando. Utilizando a linguagem natural torna-se mais simples o questionamento de uma determinada área, já que não há necessidade de se saber corretamente a implementação do sistema, o que ele irá buscar, como por exemplo, em uma consulta a um banco de dados, o usuário não precisa saber o que são tabelas e nem como elas buscam as informações, e nem o funcionamento de um banco de dados, ele apenas deseja que o resultado da pesquisa seja mostrado de forma simples e objetiva. (SILVA et al., 2007);
- g) Simplificação de Texto: Ou sumarização é a área de processamento em PLN que elabora/gera resumos (sumários) a partir de textos completos. São sistematizações feitas a partir da extração de palavras chaves identificadas no texto original. (PARDO, 2008);
- h) Correção Ortográfica: Sistemas de correção ortográfica (do inglês, spelling checker systems) processam um texto em uma dada língua natural com os objetivos de identificar os erros cometidos quanto à ortografia (palavras que não constam do léxico dessa língua ou usadas em contexto impróprio) e sugerir alternativas prováveis e ortograficamente corretas a cada erro identificado. (FELIPPO; SILVA, 2008).

4.3 SISTEMAS DE PROCESSAMENTO DE LINGUAGEM NATURAL (SPLN)

Os Sistemas de Processamento de Linguagem Natural (SPLN) são programas capazes de interpretar e/ou gerar informação fornecida em linguagem natural. (SILVA et al., 2007).

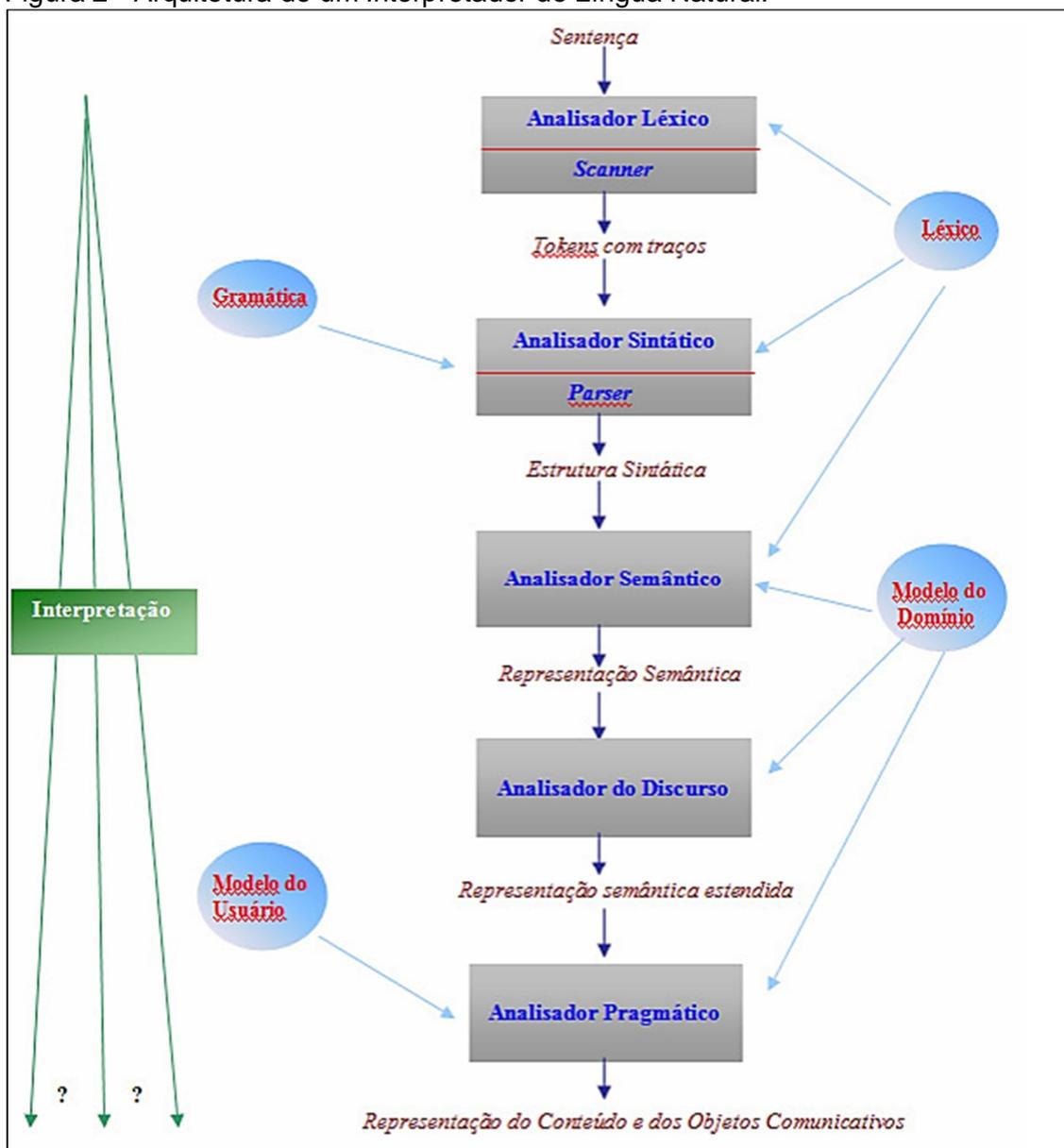
Como abordado no tópico 4.1, a tradução automática é um tipo de SPLN que abrange tanto a geração quanto a interpretação de Linguagem Natural. Um sistema com a finalidade de traduzir uma sentença de uma língua origem para uma destino, precisaria resumidamente ser capaz de reconhecer cada uma das palavras, analisar sintaticamente e semanticamente a sentença, extrair e mapear o significado em uma representação adequada e transformá-la em uma sentença da língua destino. (SILVA et al., 2007).

Como é o caso desta pesquisa, vários sistemas de PLN possuem apenas uma dessas funcionalidades, ou a geração ou a interpretação de língua natural. A seguir detalha-se o desenvolvimento de sistemas de Interpretação de Linguagem Natural.

4.3.1 Arquitetura de um Sistema de Interpretação de Linguagem Natural

A arquitetura de um sistema de interpretação de linguagem natural é composta por módulos de processamento e pelos recursos necessários para que esse processamento ocorra. Essa estrutura está representada pela Figura 2. Os módulos de processamento estão representados pelos retângulos, os recursos, pelas elipses. (SILVA et al., 2007; ROSA, 2011).

Figura 2 - Arquitetura de um Interpretador de Língua Natural.



Fonte: Silva et al. (2007).

Cada componente da arquitetura tem uma função dentro do interpretador e são responsáveis por interpretar a linguagem natural de entrada.

- Analísador Léxico ou Scanner: Responsável pela identificação e separação dos componentes mais importantes da sentença. Abrange as palavras, símbolos de pontuação, traços gramaticais e/ou semânticos de cada sentença, com base em consultas ao léxico. (SILVA et al., 2007);

- b) Analisador Sintático ou Parser: Responsável por construir ou recuperar uma estrutura sintática válida para a sentença, perdida na etapa anterior. Essa construção é feita através de uma representação da gramática da língua natural em questão. Adota-se uma gramática parcial da língua natural, contemplando apenas as construções de interesse da aplicação. Existem várias técnicas utilizadas para realizar este parsing, desde formalismos mais simples que são mais eficientes e menos abrangentes até os mais completos, abrangentes e pouco eficientes. A estrutura sintática gerada pelo parser varia de acordo com o formalismo e gramática escolhida. (SILVA et al., 2007);
- c) Analisador Semântico: Sempre que a aplicação exigir algum tipo de interpretação, este módulo será acionado. Nesse caso, é necessário conhecimento específico do domínio, para garantir interpretação correta dos termos da sentença. A estrutura semântica expressa o relacionamento dos termos a nível de significado, podendo ser representada funcionalmente pelos componentes semânticos expressos pelos componentes sentenciais. (SILVA et al., 2007);
- d) Analisador do Discurso: A análise discursiva se dá quando o significado da sentença pode depender das sentenças que a antecedem e influenciar nas que se seguem. Nestes textos multi-sentenciais, são comumente utilizadas referências anafóricas por meio de pronomes, como: “ele”, “ela”, “este” e “aquela”. Ou por meio de sinônimos, como: “a menina” referindo-se a “Amélia”. O analisador de discurso trata exatamente deste tipo de inter-relacionamento, assumindo maior responsabilidade a medida que a complexidade entre os componentes sentenciais aumenta. (SILVA et al., 2007);
- e) Analisador Pragmático: O Analisador Pragmático é responsável por verificar se existem aspectos pragmáticos da comunicação embutidos na sentença. Nem sempre o caráter interrogativo de uma sentença expressa exatamente o caráter de solicitação de uma resposta. A sentença “Você sabe que horas são?” pode ser interpretada de duas formas: uma solicitação para que as horas sejam informadas, ou uma repreensão por um atraso ocorrido. Diferenças de interpretação desse

tipo, podem ocasionar problemas na interpretação do discurso. (SILVA et al., 2007).

5 MINERAÇÃO DE DADOS

De forma geral, a Mineração de Dados ou Data Mining pode ser conceituada como a descoberta e análise inteligente de informações úteis da Web. (COOLEY, 1997).

Todos os tipos de textos que compõem o dia a dia de organizações e pessoas são produzidos e armazenados em meios digitais. Além de todos os conteúdos produzidos profissionalmente por empresas, os usuários passaram a compartilhar na web seus conhecimentos, críticas, opiniões e vincular esses conteúdos a sites, blogs, redes sociais, fóruns, bate papos, dentre outros. Desta forma, usuários tendem a postar seus comentários sobre pessoas, organizações, serviços, produtos e marcas, alimentando ainda mais esse vasto banco de informações da World Wide Web. (GUEDES; AFONSO; MAGALHÃES, 2010).

Sendo assim, torna-se evidente a dificuldade em filtrar e tirar informações relevantes dessa imensa massa de dados. Para usuários finais, é de grande interesse saber quais as demais opiniões sobre um produto que deseja adquirir. Para organizações, saber o que os consumidores pensam sobre sua marca e produtos é uma grande vantagem competitiva. Em virtude desse grande volume de dados eletrônicos disponíveis na internet e a necessidade de obter informações relevantes a partir deles, torna-se necessário o uso de técnicas de extração de conhecimento automáticas e eficientes, com o objetivo de recuperar e minerar conhecimentos úteis da web e apresentá-los ao usuário em uma leitura objetiva e conclusiva, facilitando a interpretação e tomada de decisão. (GUEDES; AFONSO; MAGALHÃES, 2010).

Existem três frentes que categorizam a Mineração de Dados na Web, a Mineração de Uso, a Mineração de Estrutura e a Mineração de Conteúdo. A mineração de uso aborda a mineração das informações de uso da Web, são as informações sobre como o usuário interage com a Web. Nessa categoria são tratadas questões como personalização, interfaces adaptativas e aprendizado de perfis de usuários. A mineração de estrutura aborda a mineração das informações contidas entre os documentos da Web. Os documentos da Web se relacionam basicamente através de vínculos de hipertexto, e esses vínculos escondem informações valiosas não só sobre a topologia da Web, mas também sobre como os documentos se relacionam. A mineração de conteúdo aborda a mineração dos

dados contidos dentro dos documentos da Web. A grande quantidade de formatos que os dados podem assumir (textos comuns, páginas HTML, imagens, áudio, vídeo, etc.) acaba dirigindo as técnicas de mineração a serem utilizadas. (MARINHO; GIRARDI, 2005).

Esta última se estende para a Mineração de Opinião ou Opinion Mining, responsável por minerar e classificar opiniões, assunto do próximo tópico e de grande importância para esta pesquisa.

5.1 MINERAÇÃO DE OPINIÃO

A mineração de opinião, ou Opinion Mining, também conhecida como análise de sentimentos, pode ser definida como a técnica que avalia um conteúdo subjetivo emitido em linguagem natural e descobre o sentimento que é transmitido. Geralmente associado à classificação binária entre sentimentos positivos e negativos, o termo é usado de uma forma mais abrangente para significar o tratamento computacional de opinião, sentimento e subjetividade em textos. (PANG; LEE, 2002). Com o advento da web como fonte de informações, grande parte dos usuários tem buscado nela textos que forneçam esse tipo de informação desejada, opiniões sobre alguma entidade de interesse como um produto específico, uma empresa, um lugar, uma pessoa, dentre outros. O objetivo principal é permitir que um usuário obtenha uma visão geral sobre o que outros pensam sobre o produto, sem precisar localizar e ler cada opinião feita na web. Para atingir esse objetivo, a mineração de opinião é dividida em três grandes etapas: coleta de conteúdo, classificação e sumarização dos resultados.

5.1.1 Coleta de Conteúdo

Etapa na qual é feita uma busca em fontes diversas, tais como artigos em sites, comentários em mídias sociais, anúncios, documentos dentre outras. É importante a utilização de técnicas avançadas de busca, visando identificar se o conteúdo encontrado trata-se de uma opinião ou um fato. Fatos por si só devem ser descartados, porém opiniões expressas em fatos devem ser mantidas. (BECKER; TUMITAN, 2005).

5.1.2 Classificação

A classificação é a etapa mais importante do processo e é nela que a polaridade ou orientação da opinião é definida. Esta etapa determina se uma opinião é positiva, negativa ou neutra. Na classificação ou análise de sentimentos, são as palavras opinativas que têm a maior importância, pois, através delas, é possível determinar o sentimento expresso pelo autor. Exemplos de palavras opinativas: bom, legal, ótimo, ruim, péssimo etc. (BECKER; TUMITAN, 2005).

5.1.3 Sumarização dos Resultados

Etapa focada na apresentação dos resultados, que podem ser de forma textual, ou gráfica. A melhor forma de representar os resultados é a gráfica, pois facilita a visualização e entendimento dos resultados sumarizados em totais e dados estatísticos. (BECKER; TUMITAN, 2005).

6 TRABALHOS CORRELATOS

Embora as áreas de Processamento de Linguagem Natural e Mineração de Opinião sejam recentes, existem muitas pesquisas e informações disponíveis. A cada dia surgem novos artigos científicos, novas metodologias para alcançar os resultados, novas ferramentas de classificação de sentimentos, novos bancos de dados para utilização, novos desafios e necessidades. Isso prova que as áreas estão aquecidas e em crescente expansão.

Com a gama de pesquisas disponíveis, a metodologia que será utilizada nesta pesquisa será direcionada de acordo com resultados obtidos anteriormente, evitando que os mesmos erros sejam cometidos e que seja dada mais atenção aos pontos críticos.

De acordo com outras pesquisas, um ponto crítico é a utilização de ferramentas para tradução de sentenças, procedimento que se faz presente neste trabalho e que pode causar algum tipo de perda ao sentido ou significado das opiniões. Tendo esse esclarecimento, essa pesquisa será direcionada para utilizar uma ferramenta que não cause esses danos ao processo.

Uma das pesquisas analisadas foi a intitulada: “Protótipo para Mineração de Opinião em Redes Sociais: Estudo de Casos selecionados usando o Twitter”, de autoria de Leandro Matioli Santos, que aborda a mineração da opinião nas redes sociais, estudando se é possível aplicar a técnica nesse tipo de mídia, quais os desafios e dificuldades, se os resultados são satisfatórios e relevantes. (SANTOS, 2010).

Outra pesquisa analisada foi a intitulada: “Mineração de Opiniões aplicada a mídias sociais”, de Marlo Vieira dos Santos e Souza, que analisa o cenário das empresas nas mídias sociais, a inteligência competitiva, o mercado capitalista e suas tendências nessas mídias, opiniões sobre produtos, marcas, entidades, etc. (SOUZA, 2012).

Dessa forma, a intenção deste trabalho é contribuir com informações, conclusões e análises relevantes para essas áreas, direcionando para a exploração de opiniões nas mídias sociais envolvendo clientes e organizações.

7 METODOLOGIA

A metodologia de desenvolvimento deste trabalho consistiu na definição e concretização de três etapas: obtenção do corpus de comentários, arquitetura do sistema proposto e desenvolvimento.

7.1 CORPUS DE COMENTÁRIOS

Para o presente trabalho ser realizado, foi necessária a construção de um corpus de comentários que serviu como entrada de dados para que o sistema realizasse a mineração da opinião, conforme técnica explanada no tópico anterior. Esse corpus foi formado através da rede social Facebook, utilizando fanpages de organizações reais para coleta dos dados, tais como Sony¹, Samsung², Panasonic³, Brastemp⁴ e Electrolux⁵. Foram consultadas as fanpages brasileiras das marcas e puderam-se obter tanto comentários positivos quanto negativos. As Figuras 3 e 4 mostram, por exemplo, um comentário positivo e um negativo, retirados da fanpage da Sony Brasil.

Figura 3 – Opinião positiva.



Fonte: Fanpage da Sony Brasil.

A opinião da Figura 3 é um comentário de uma publicação da Sony e expressa um sentimento positivo quanto aos produtos da marca. Nota-se expressões como “espetacular” e “valeu Sony”, que revelam o sentimento passado pelo autor.

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

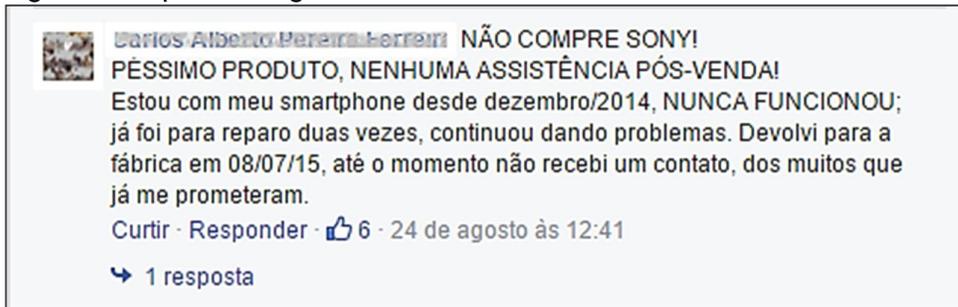
² <https://www.facebook.com/SamsungBrasil/>

³ <https://www.facebook.com/panasonic.br/>

⁴ <https://www.facebook.com/assimumabrastemp/>

⁵ <https://www.facebook.com/ElectroluxBrasil/>

Figura 4 - Opinião Negativa.



Fonte: Fanpage da Sony Brasil.

Já a opinião da Figura 4, expressa um sentimento negativo sobre a marca. Expressões como “PÉSSIMO PRODUTO” e “NÃO COMPRE”, revelam o sentimento transmitido pelo autor.

A rede social Facebook foi escolhida devido a sua popularidade, alto número de usuários conectados e, principalmente, a variedade de tipos de usuários, pessoas físicas e jurídicas, característica importante para esta investigação por promover a interação entre empresas e consumidores.

O corpus construído foi inserido em uma fanpage fictícia, que simulou uma fanpage real de uma organização. Para que o sistema analise e classifique estas opiniões, basta acessá-lo, informar o login e senha do usuário no Facebook e escolher de qual fanpage deseja baixar e classificar opiniões.

Só serão consideradas as publicações feitas dentro da fanpage da organização. Se a publicação for feita por um consumidor, ou seja, outro usuário do Facebook, a própria postagem será classificada. Caso a publicação tenha sido feita pela própria fanpage, apenas os comentários desta serão classificados. Essa distinção foi necessária devido à grande quantidade de usuários que expressam suas opiniões em publicações feitas pela própria empresa.

7.1.1 ARQUITETURA DO SISTEMA

O sistema foi dividido em três fases:

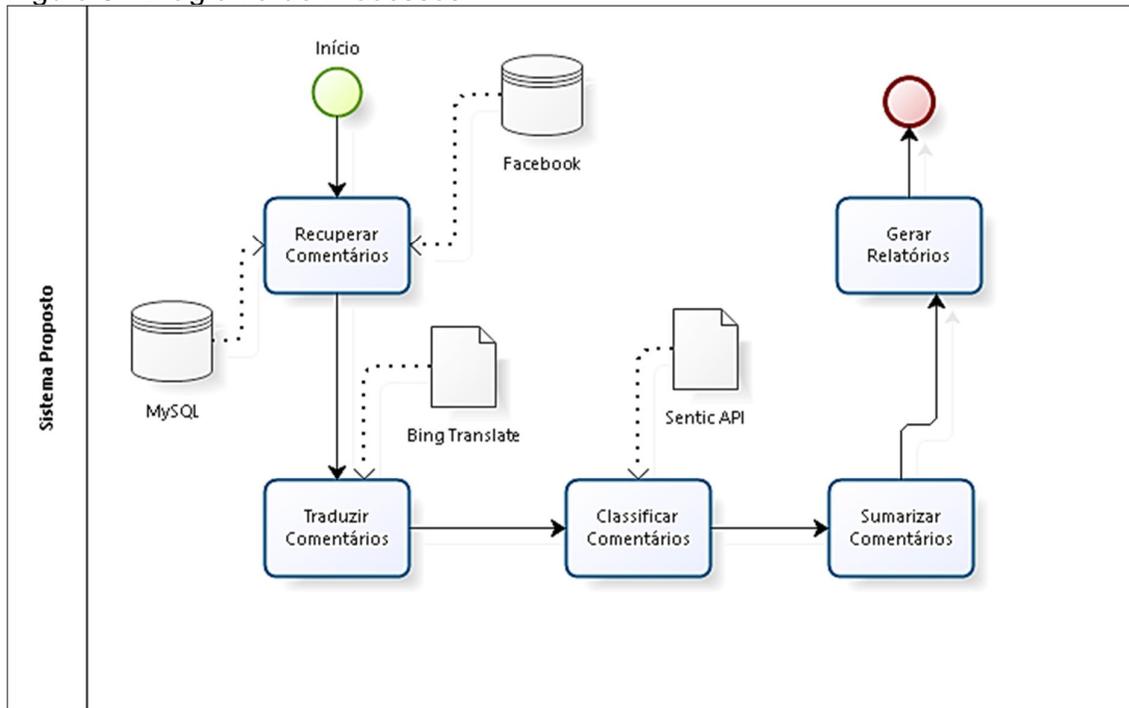
- a) **Primeira Fase:** Responsável por acessar a *fanpage* da empresa no Facebook, baixar os comentários publicados e salvá-los no banco de dados MySQL. O acesso a fanpage foi feito através de login e senha

da conta no Facebook. Esses comentários tiveram suas polaridades classificadas na segunda fase;

- b) **Segunda Fase:** Responsável por acessar no banco MySQL, os comentários baixados na primeira fase e classifica-los de acordo com sua polaridade. Os comentários/opiniões podem assumir caráter positivo, negativo ou neutro;
- Opinião de caráter positivo: Opiniões de caráter positivo são aquelas em que aparecem recomendações ou elogios sobre a marca ou produto. Transmitem o sentimento de aprovação ao receptor da opinião;
 - Opinião de caráter negativo: Opiniões de caráter negativo são aquelas em que aparecem reclamações ou não recomendações sobre a marca ou produto. Transmitem o sentimento de reprovação ao receptor da opinião;
 - Opinião de caráter neutro: Opiniões de caráter neutro são aquelas que não transmitem sentimentos positivos ou negativos ao receptor da opinião.
- c) **Terceira Fase:** Responsável por recuperar as sentenças classificadas na segunda fase e montar um relatório conclusivo. Esse relatório será a visão final para o usuário, que poderá tomar decisões e estabelecer planos de ação mediante ao cenário apresentado.

O funcionamento do sistema é ilustrado na Figura 5, através do diagrama de processos.

Figura 5 - Diagrama de Processos.



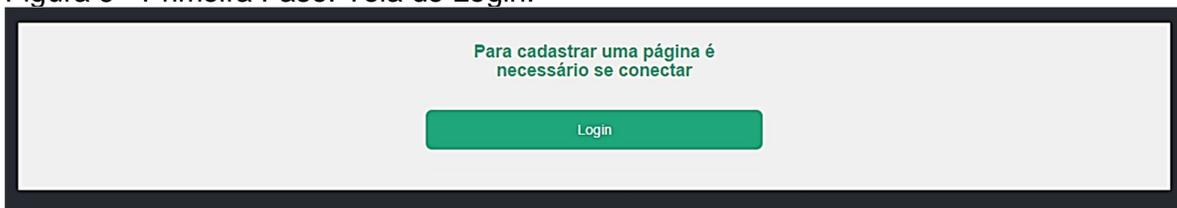
Fonte: Elaborada pela autora.

7.2 DESENVOLVIMENTO

A primeira fase contempla a interface do sistema, a comunicação com o Facebook e o download das opiniões postadas na fanpage da empresa.

No primeiro acesso, o usuário visualizará a tela de login, onde utilizará sua conta do Facebook para logar, conforme mostrado na Figura 6. Em seguida, serão exibidas em uma lista, as fanpages vinculadas à conta. O usuário deverá escolher qual fanpage deseja utilizar e indicar qual será a padrão. Conforme ilustrado na Figura 7, serão então carregadas as publicações feitas na fanpage escolhida.

Figura 6 - Primeira Fase: Tela de Login.



Fonte: Elaborada pela autora.

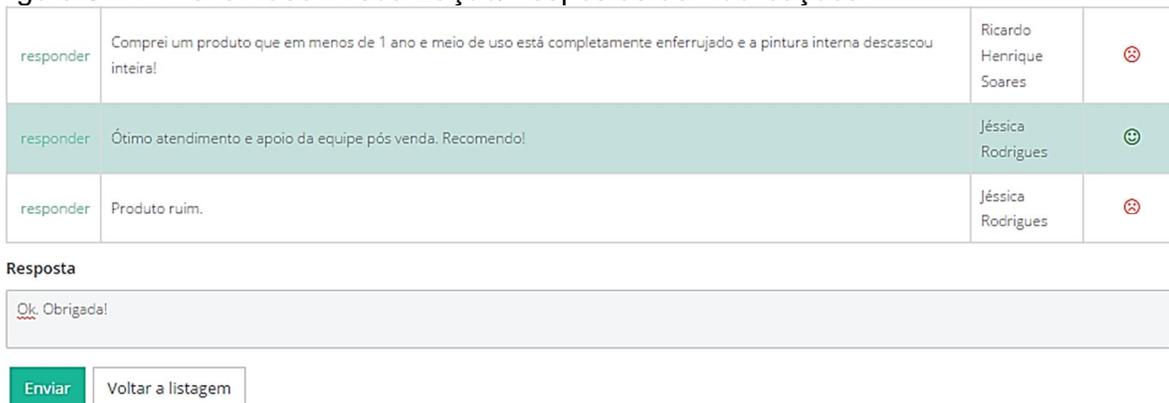
Figura 7 - Primeira Fase: Listagem de publicações feitas na fanpage.



Fonte: Elaborada pela autora.

O usuário pode então escolher uma publicação para visualizar e responder, conforme ilustrado na Figura 8.

Figura 8 - Primeira Fase: Visualização/Respostas de Publicações.



Fonte: Elaborada pela autora.

Ao enviar a resposta, esta é publicada como um comentário da postagem inicial, conforme ilustra a Figura 9.

Figura 9 - Primeira Fase: Postagem na Fanpage através do sistema.



Fonte: Elaborada pela autora.

Conforme mostrado nas Figuras 7 e 8, o sistema permite o gerenciamento das publicações feitas nas fanpages de uma conta da rede Facebook. A comunicação com o Facebook ocorre através da Application Programming Interface (API) disponibilizada pelo próprio Facebook à desenvolvedores de software. Como a API foi desenvolvida na linguagem de programação PHP, a mesma linguagem foi utilizada para o desenvolvimento da primeira fase deste sistema. O MySQL foi utilizado para salvar as opiniões postadas na fanpage, eliminando a necessidade de várias consultas ao Facebook, o que deixaria o processamento mais lento. Esse banco foi escolhido devido a sua facilidade de utilização e sua grande compatibilidade com as linguagens de programação utilizadas neste sistema (PHP na primeira fase e Java nas segunda e terceira fases).

Na segunda fase, cada opinião recuperada é classificada entre positiva, negativa ou neutra, de acordo com o sentimento transmitido. A Figura 10 mostra onde é exibida essa informação.

Figura 10 - Segunda Fase: Opiniões classificadas.

Visualizar/Responder Post

Quando o assunto é oferecer ainda mais comodidade para você, a Samsung French Door Sparkling não hesita: água com gás na porta do seu refrigerador e uma dose de inovação e praticidade! [responder]

Comentários			
	Mensagem	Usuário	Polaridade
responder	Empresa confiável e de boa reputação. Voltaria a fazer negócio.	Jéssica Rodrigues	😊
responder	PROPAGANDA ENGANOSA!!! Consultem o site RECLAME AQUI!	Jéssica Rodrigues	😡
responder	Beta melhor em tudo que faz,e parabéns pra nós .	Jéssica Rodrigues	😊
responder	Decepcionada com a Beta, as propagandas são lindas e maravilhosas, tenho 2 meses e pouco de uso do meu aparelho celular e já apresentou defeito.	Jéssica Rodrigues	😊
responder	TOTAL DESCASO COM O CLIENTE, MANDO PRA ASSISTENCIA E ME DEVOLVEM O PRODUTO ESTRAGADO E AGORA NÃO QUEREM REPARAR O ERRO!!!!!!	Jéssica Rodrigues	😡
responder	OLÁ BETA! GOSTARIA DE DEIXAR BEM CLARO AQUI PARA OUTRAS PESSOAS COMO O SEUS SERVIÇOS DE ATENDIMENTO AO CONSUMIDOR É INOPERANTE E DESQUALIFICANTE COM SEUS CLIENTES.	Jéssica Rodrigues	😊

Fonte: Elaborada pela autora.

A classificação se dá através da API (Sentic API) do SenticNet⁶, software criado pelo Sentic Team, grupo multidisciplinar de pesquisa dentro da NTU School of Computer Science. O objetivo era fazer com que a informação conceitual transmitida via linguagem natural fosse facilmente acessível às máquinas.

A Sentic API disponibiliza estruturas no formato XML onde armazena grande quantidade de expressões na língua inglesa. As expressões são classificadas com uma polaridade que varia de -1 a 1, onde 1 é o mais positivo possível e -1 o mais negativo possível. O zero (0) classifica a expressão como neutra. Por exemplo, a expressão "friday night", expressão que traz sentimento positivo, possui polaridade +0.728. Já a palavra "monday", dia da semana não tão popular, tem valor -0.847.

A fim de obter uma melhor performance junto a Sentic API, a sentença é dividida em partes e passada dessa forma. Cada parte recebe um score de polaridade e ao final é calculada uma média das pontuações. A média indica a polaridade da sentença como um todo.

Como a Sentic API possui apenas expressões em inglês e os comentários coletados são em português, foi necessário utilizar uma ferramenta de tradução. Por ser gratuita, foi adotada a ferramenta Bing Translator, do sistema de buscas Bing da

⁶ Disponível em: <<http://sentic.net/>>.

Microsoft. Basicamente, esta ferramenta é uma API que recebe a sentença e a língua para qual se deseja traduzir e a retorna traduzida. A eficácia deste recurso foi validada no momento do desenvolvimento desta fase do sistema, e apresentou resultado satisfatório para a tradução, não havendo perdas no sentido das sentenças.

Para o desenvolvimento desta fase, foi adotada a linguagem Java devido a familiaridade da autora com a linguagem.

Com as sentenças devidamente classificadas, a terceira fase do sistema se encarrega de tratar essas informações e montar uma visão conclusiva sobre o cenário da organização na rede social Facebook.

As opiniões classificadas são agrupadas e exibidas em um relatório através de gráficos, possibilitando a visualização da intensidade dos sentimentos. Foi desenvolvida também uma visão de opinião por consumidor, permitindo a organização saber quantas vezes cada consumidor comentou em sua página. Por fim, uma visão mensal e anual trazendo os sentimentos do período.

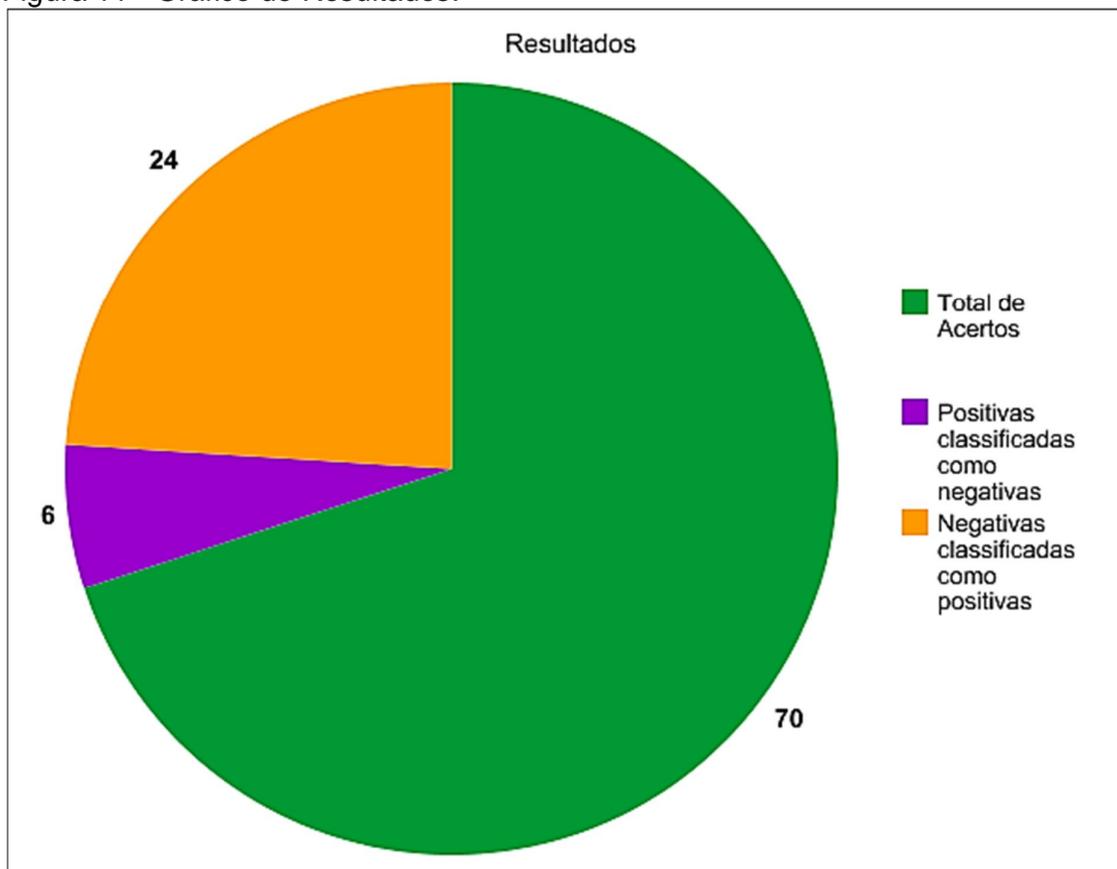
Esta fase também foi desenvolvida com a linguagem Java pelo mesmo motivo da segunda fase.

8 RESULTADOS

O sistema proposto foi submetido a testes de precisão para definição de sua eficiência. A precisão é medida, considerando o número de acertos do sistema, ou seja, se o sistema classifica corretamente as opiniões, previamente rotuladas por juízes humanos, em positivas ou negativas.

No treinamento, foram cadastradas na fanpage uma amostra de cem opiniões, cinquenta positivas e cinquenta negativas. A Figura 11 mostra em forma gráfica o resultado obtido.

Figura 11 - Gráfico de Resultados.



Fonte: Elaborada pela autora.

Foram computados um total de 70% de acertos das cem opiniões classificadas, número expressivo que mostra o bom desempenho do sistema. Da amostra, seis classificações foram classificadas erroneamente como negativas e vinte e quatro como positivas.

Alguns desses resultados inesperados serão explicados a seguir.

Figura 12 - Precisão do classificador.

Opinião	Juízo Humano	Sistema Proposto
Gostei muito do produto, mas a câmera frontal dele não é 100%.	Positivo	Negativo
Celular espetacular, mas falha num requisito fundamental, durabilidade.	Negativo	Positivo
É um excelente Celular, navegação internet excelente, tela nitida, câmera ótima, não trava, especialmente na hora de fazer selfs com o guia é um diferencial e tanto.	Positivo	Positivo
Tenho um fone sony há mais de um ano e sem dúvidas e muito show.	Positivo	Positivo
Sony é um lixo não comprem nada, é uma ilusão.	Negativo	Negativo
Quero q vcs cubram a garantia do meu ps4 a cada dia se tornando uma empresa horrível com o trato com o cliente	Negativo	Negativo

Fonte: Elaborada pela autora.

Na primeira opinião da Figura 12, o termo “produto” possui uma polaridade positiva de 0.154 na Sentic API. Já os termos “câmera” e “frontal” possuem respectivamente polaridades negativas de -0.026 e -0.674, valores que definem o sentimento negativo da opinião.

Por juízo humano, a polaridade desta sentença foi definida como positiva, porém, a opinião carrega a preposição adversativa “mas”, que cria um contraste de ideias na frase. Com este contraste, podemos enxergar duas opiniões e sentimentos na opinião: antes do “mas”, “Gostei muito do produto”, com um sentimento positivo e depois do “mas”, “a câmera frontal dele não é 100%.”, com um sentimento negativo.

Na segunda opinião da Figura 10, o motivo do resultado não esperado é o mesmo da primeira opinião.

Logo, podemos concluir que tratam-se de sentenças com um nível de dificuldade alto de identificação de um único sentimento.

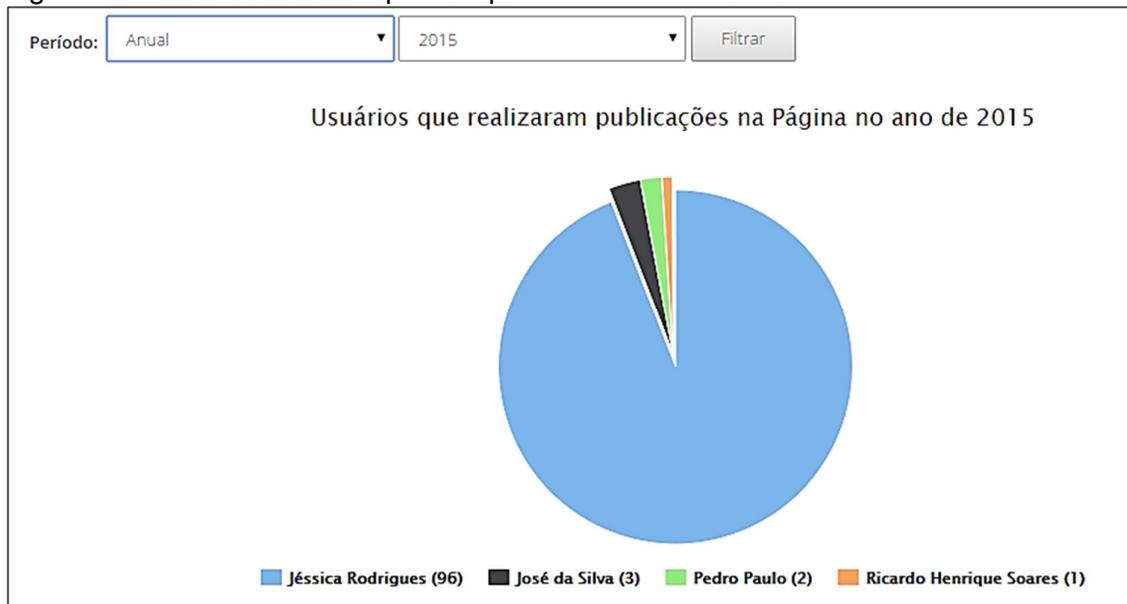
As demais opiniões apresentadas, mostram casos de sucesso na classificação de acordo com juízo humano.

8.1 VISÕES APRESENTADAS PELO SISTEMA

Conforme apresentado na Metodologia do presente trabalho, a terceira fase do sistema proposto tinha como objetivo a criação de visões gráficas que auxiliarão

na tomada de decisão da organização. A Figura 13 mostra os resultados do primeiro relatório gráfico disponível.

Figura 13 - Quantidade de Opiniões por consumidor no ano.

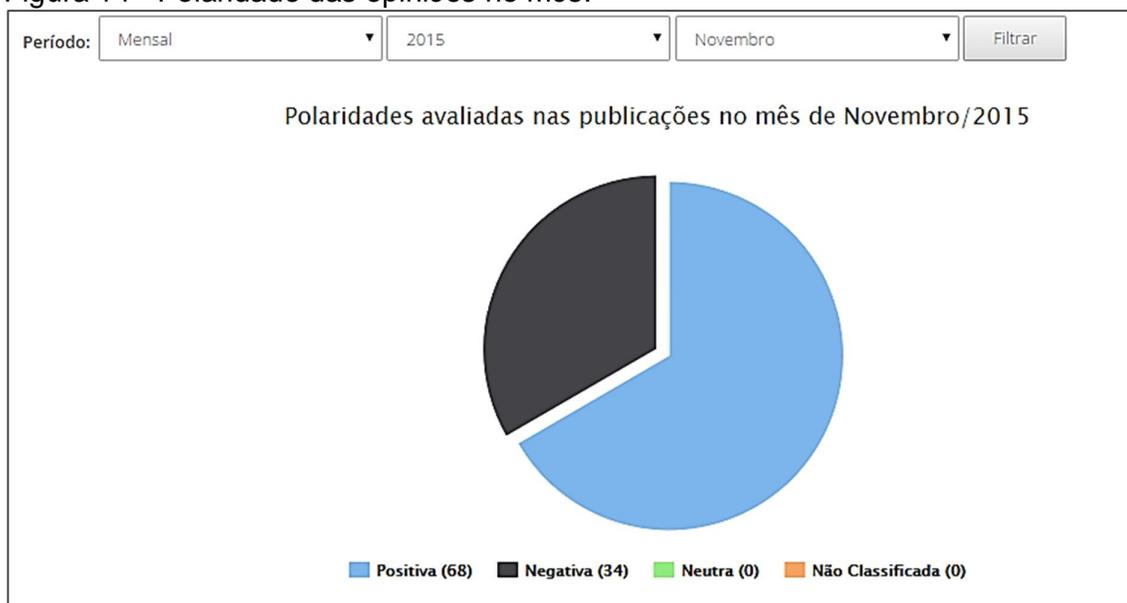


Fonte: Elaborada pela autora.

O gráfico revela que no ano de 2015 houve cento e duas opiniões postadas na fanpage. Essa visão é importante para que a organização tenha a informação de quantos consumidores diferentes estão opinando sobre seus serviços ou produtos, quais são os que mais opinam e se essas opiniões acontecem em um curto período de tempo entre uma e outra.

A Figura 14 mostra os resultados do segundo relatório.

Figura 14 - Polaridade das opiniões no mês.

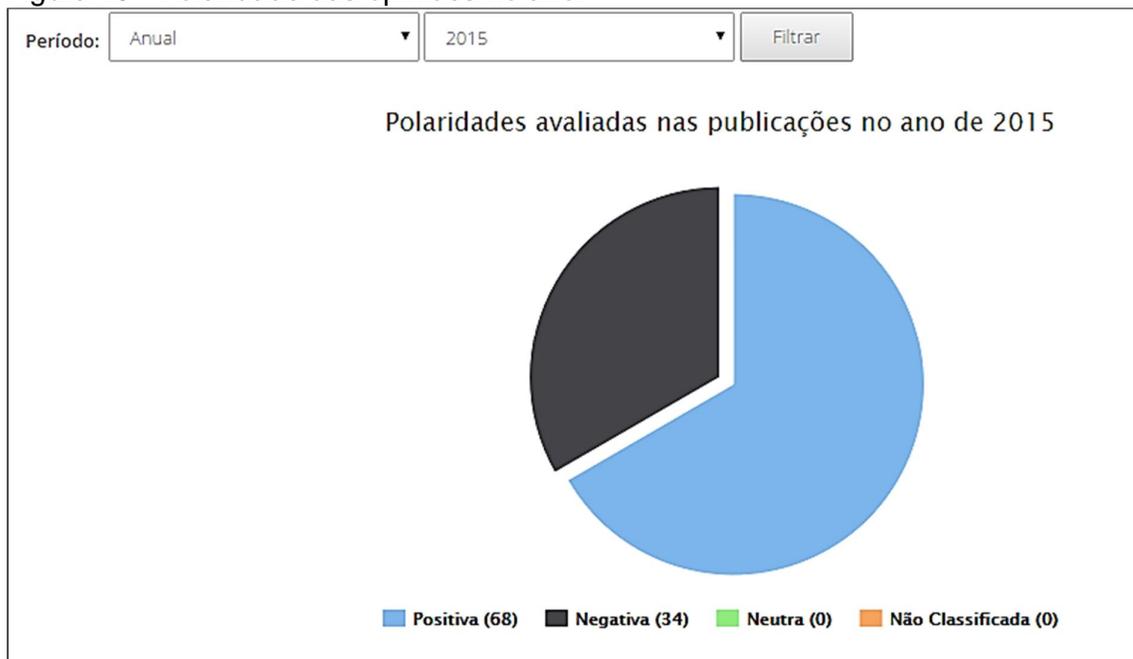


Fonte: Elaborada pela autora.

O gráfico mostra a quantidade de opiniões positivas, negativas e neutras no período, que pode ser mensal ou anual. Essa visão é a mais importante por mostrar à organização qual é o sentimento predominante de seus consumidores com relação à marca. A partir deste gráfico, a empresa será capaz de saber a intensidade dos sentimentos de seus consumidores ao decorrer dos meses, tendo a possibilidade de concluir se uma alta nos sentimentos negativos ocorreram devido ao lançamento de um determinado produto ou serviço, ou se sentimentos positivos aumentaram devido à uma campanha publicitária lançada, podendo evitar ou repetir as estratégias na rede Facebook.

A Figura 15 mostra a visão do gráfico acima, porém com filtro de período anual, possibilitando à organização enxergar seus resultados durante os anos.

Figura 15 - Polaridade das opiniões no ano.



Fonte: Elaborada pela autora.

9 CONSIDERAÇÕES FINAIS

O presente trabalho teve como objetivo a pesquisa e implementação da mineração de opinião na rede social Facebook. A implementação se deu através de um software que obteve um resultado satisfatório diante do esperado, sendo capaz de classificar corretamente 70% das opiniões da amostra apresentada. Apesar das opiniões do corpus montado serem reais, novos testes podem ser realizados utilizando uma fanpage de uma organização real, com uma marca expressiva de comentários. Caso os resultados continuem satisfatórios nesse cenário, a ferramenta poderia começar a ser utilizada por empresas.

Porém, mesmo com um bom resultado, existem alguns pontos que podem ser melhorados, como, por exemplo, a velocidade do processamento da classificação. Atualmente, esse processamento é um tanto oneroso em questão de tempo, e esse fator é hoje muito escasso para as empresas. Outro ponto a ser melhorado é a utilização da classificação através de janelas, o que corresponderia ao envio de mais de uma palavra (formando expressões) para obtenção da polaridade na Sentic API. Atualmente, o sistema proposto envia uma palavra por vez. A Sentic API oferece suporte para esta melhoria, possuindo expressões como “a lot of books”, que significa “muitos livros” e tem polaridade positiva de 0.047. Outro exemplo é “a lot of noise” que significa “muito barulho”, com polaridade negativa de -0.619. Essa implementação provavelmente traria uma classificação ainda mais precisa do sentimento, pois carrega um conhecimento mais específico.

Apesar das limitações citadas, o foco particular do processamento da língua portuguesa, está entre as principais contribuições deste trabalho, não só para a validação de um método e criação de um sistema de análise de sentimentos, mas principalmente porque há poucos recursos para esta língua, uma vez que a maior parte das pesquisas envolve o idioma inglês.

Desta forma, conclui-se que o presente trabalho contribuiu com conhecimento para o segmento de Análise de Sentimentos da Inteligência Artificial e Ciência da Computação, explorando mais uma forma de implementar a interpretação de linguagem natural e identificação de sentimentos através de um software. Além disso, apesar de ainda precisar de melhorias, o software pode ser utilizado e aliado das organizações, provendo informações sobre as marcas, através da classificação das opiniões dos consumidores.

REFERÊNCIAS

BECKER, K.; TUMITAN, D. Introdução à Mineração de Opiniões: Conceitos, Aplicações e Desafios. **UFRGS**, 2005. Disponível em: <http://www.inf.ufrgs.br/~kbecker/lib/exe/fetch.php?media=minicursosbbd_versaosubmetida.pdf>. Acesso em: 10 maio 2015.

COOLEY, R. Web mining: information and pattern Discovery on the World Wide Web, Proceedings of the 9th IEEE International Conference on Tools with Artificial Intelligence. **IEEEExplore**, 1997. Disponível em: <<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=632228>>. Acesso em: 10 maio 2015.

DOURADO, D. et al. **Mídias Sociais: perspectivas, tendências e reflexões**. Bahia: PaperCliq, 2010.

FELIPPO, A. D. et al. Uma introdução à Engenharia do Conhecimento Linguístico. **Revista de Letras da Universidade Católica de Brasília**, Brasília-DF, v. 1, n. 2, ano 1, p. 57-72, nov. 2008. Disponível em: <<http://portalrevistas.ucb.br/index.php/RL/article/viewFile/786/825>>. Acesso em: 4 maio 2015.

GUEDES, R.; AFONSO, D.; MAGALHÃES, L. H. de. Mineração de opiniões de usuários na busca de conhecimento. **Vianna Sapiens**, 2010. Disponível em: <http://www.viannajunior.edu.br/files/uploads/20131001_141137.pdf>. Acesso em: 4 maio 2015.

GONGORA, A. D. O que é Inteligência Artificial? **UFSC**, 2007. Disponível em: <<http://www.egov.ufsc.br/portal/sites/default/files/anexos/6515-6514-1-PB.pdf>>. Acesso em: 26 abr. 2015.

IBM BRASIL. Computação Cognitiva: 5 Futuras Inovações Tecnológicas. **Youtube**, 2013. Disponível em: <<https://www.youtube.com/watch?v=X2l1-9yMhzM>>. Acesso em: 26 abr. 2015.

LEVY, P. **A inteligência coletiva: por uma antropologia do ciberespaço**. São Paulo: Loyola, 2000.

MARINHO, L. B.; GIRARDI, R. Mineração na Web. **Research Gate**, 2005. Disponível em: <http://www.researchgate.net/profile/Leandro_Marinho/publication/228369738_Minerao_na_Web/links/0deec539852e70896c000000.pdf>. Acesso em: 10 maio 2015.

OLIVEIRA, A. C. **Inteligência competitiva na internet**. Rio de Janeiro: Brasport, 2006.

PANG, B.; LEE, L.; VAITHYANATHAN, S. Thumbs up? Sentiment Classification using Machine Learning Techniques. **ACM DIGITAL LIBRARY**, 2002. Disponível em: <<http://dl.acm.org/citation.cfm?id=1118704>>. Acesso em: 04 jun. 2015.

PARDO, T. A. S. Sumarização Automática: Principais Conceitos e Sistemas para o Português Brasileiro. **NILC-USP**, 2008. Disponível em: <<http://www.icmc.usp.br/~tasparado/NILCTR0804-Pardo.pdf>> Acesso em: 26 abr. 2015.

PORTER, M. **Estratégia competitiva**. Rio de Janeiro: Campus, 1986.

RICH, E. **Inteligência Artificial**. São Paulo: McGRALL-HILL, 1988.

ROHR, A. Computador convence juízes de que é garoto de 13 anos em 'teste de Turing'. **Globo.com**, 2014. Disponível em: <<http://g1.globo.com/tecnologia/noticia/2014/06/computador-convence-juizes-que-e-garoto-de-13-anos-em-teste-de-turing.html>>. Acesso em: 26 abr. 2015.

ROSA, J. L. G. **Fundamentos da Inteligência Artificial**. Rio de Janeiro: Gen-LTC, 2011.

RUSSEL, S. J.; NORVIG, P. **Inteligência Artificial**. Rio de Janeiro: Elsevier, 2004.

SANTOS, L. M. Protótipo para Mineração de Opinião em Redes Sociais: Estudo de Casos selecionados usando o Twitter. **UFLA**, 2010. Disponível em: <<http://www.bcc.ufla.br/wp-content/uploads/2013/2010/LeandroMatioli.pdf>>. Acesso em: 26 abr. 2015.

SATO, P. O que é Inteligência Artificial? **Revista Escola**, 2009. Disponível em: <<http://revistaescola.abril.com.br/ciencias/fundamentos/inteligencia-artificial-onde-ela-aplicada-476528.shtml>>. Acesso em: 26 abr. 2015.

SILVA, B. C. D. da. et al. **Introdução ao Processamento das Línguas Naturais e Algumas Aplicações**. 2007. 119 f. Série de Relatórios do Núcleo Interinstitucional de Lingüística Computacional, São Carlos, 2007. Disponível em: <<http://www.letras.etc.br/ebralcn/NILCTR0710-DiasDaSilvaEtAl.pdf>>. Acesso em: 4 maio 2015.

SILVA, T. M. S. Extração de Informação para Busca Semântica na Web baseada em Ontologias. **UFSC**, 2003. Disponível em: <<https://repositorio.ufsc.br/bitstream/handle/123456789/85791/194975.pdf?sequence=1>>. Acesso em: 26 abr. 2015.

SOUZA, M. V. dos. S. Mineração de Opiniões aplicada a mídias sociais. **PUCRS**, 2012. Disponível em: <<http://meriva.pucrs.br:8080/dspace/bitstream/10923/1457/1/000448645-Texto%2bCompleto-0.pdf>>. Acesso em: 26 abr. 2015.

TAURION, C. Preparado para a era da computação cognitiva? **Cio**, 2014. Disponível em: <<http://cio.com.br/tecnologia/2014/10/07/preparado-para-a-era-da-computacao-cognitiva/>> Acesso em: 26 abr. 2015.

TURING, A. M. Computing machinery and intelligence. **UMBC**, 2002. Disponível em: <<http://www.csee.umbc.edu/courses/471/papers/turing.pdf>>. Acesso em: 26 abr. 2015.

Mineração de Opiniões Aplicada a Mídias Sociais para as Organizações

Jéssica R. Silva, Patrick P. Silva, Elvio G. Silva, Henrique P. Martins

Universidade Sagrado Coração (USC)
Caixa Postal 511 – 17.011-160 – Bauru – SP – Brasil

***Abstract.** Nowadays, the Internet is the medium in which more consumers express their views on the experience of organizations, exposing positive and negative feelings. This new behavior, allows the extraction of information that let you know how the brand is positioned in the market. This study aimed to create a software that connects the social network Facebook and ranks the opinions published. The software achieved 70% accuracy while processing a corpus of one hundred real traveler reviews, a result that reveals a good performance and indicates that there may be an ally of organizations.*

***Resumo.** Nos dias atuais, a internet é o meio em que os consumidores mais expressam suas opiniões sobre a experiência com as organizações, expondo sentimentos positivos e negativos. Esse novo comportamento, propicia a extração de informações que permitam saber como a marca está posicionada no mercado. O presente trabalho teve como objetivo a criação de um software que se conecta a rede social Facebook e classifica as opiniões publicadas. O software obteve 70% de acertos ao processar um corpus de cem opiniões reais, resultado que revela um bom desempenho e indica que pode ser um aliado das organizações.*

1. Introdução

Toda a informação produzida e disponibilizada em meios como a internet, é de difícil manejo. Antigamente, poucos tinham a capacidade de produzir informação, enquanto muitos atuavam como meros consumidores. Hoje, muitos constroem e consomem grande quantidade de informação, nos desafiando a desbravar este cenário com os mais variados recursos e metodologias para extrair disso informações relevantes e conclusivas. (LEVY, 2000).

O que uma organização deseja, é saber como sua marca está posicionada no mercado, o que ela pode melhorar baseada na opinião de seu consumidor, qual decisão tomar diante do panorama em que está inserida. (PORTER, 1986). Para isso, visto a precariedade das ferramentas atualmente disponíveis, faz-se necessário o desenvolvimento de softwares a fim de classificar a polaridade das opiniões sobre a marca, prover informações estratégicas e táticas para reduzir a incerteza associada a uma decisão e expor diretamente a atitude e pensamento de seu consumidor. (OLIVEIRA, 2006).

Neste contexto, o presente trabalho teve como objetivo a criação de um software que se conecta a rede social Facebook através da conta de uma organização, identifica o sentimento transmitido em cada opinião postada e classifica cada um entre positivo, negativo ou neutro, gerando informação relevante para a organização.

2. Processamento de linguagem natural

O objetivo do Processamento de Linguagem Natural (PLN) é fornecer aos computadores a capacidade de entender e compor textos. E "entender" um texto significa reconhecer o contexto, fazer análise sintática, semântica, léxica e morfológica, criar resumos, extrair informação, interpretar os sentidos e até aprender conceitos com os textos processados.

O PLN está voltado a três aspectos da comunicação em língua natural:

- a) som: prosódia e fonologia;

- b) estrutura: morfologia e sintaxe;
- c) significado: semântica e pragmática.

A prosódia está relacionada aos padrões de ritmo e entonação da língua. A fonologia está relacionada com o reconhecimento dos sons que compõem as palavras de uma língua. A morfologia estuda a construção das palavras, com seus radicais e afixos, que correspondem a partes estáticas e variantes das palavras, como as inflexões verbais. A sintaxe define a estrutura de uma frase, com base na forma como as palavras se relacionam nessa frase. A semântica associa significado a uma estrutura sintática, em termos dos significados das palavras que a compõem. A pragmática verifica se o significado associado à uma estrutura sintática é realmente o significado mais apropriado no contexto considerado. (ROSA, 2011).

2.1 Aplicações

O atual estágio de desenvolvimento do PLN já permite sua aplicação em áreas como: Extração de Informação, Recuperação de Informação, Tradução Automática, Geração Automática de Texto, Geração de Linguagem Natural, Simplificação de Texto, Correção ortográfica e a que diz respeito a esta pesquisa, Interpretação de Linguagem Natural. (SILVA et al., 2007).

3. Mineração de dados

De forma geral, a Mineração de Dados ou Data Mining pode ser conceituada como a descoberta e análise inteligente de informações úteis da Web. (COOLEY, 1997).

Todos os tipos de textos que compõem o dia a dia de organizações e pessoas são produzidos e armazenados em meios digitais. Além de todos os conteúdos produzidos profissionalmente por empresas, os usuários passaram a compartilhar na web seus conhecimentos, críticas, opiniões e vincular esses conteúdos a sites, blogs, redes sociais, fóruns, bate papos, dentre outros. Desta forma, usuários tendem a postar seus comentários sobre pessoas, organizações, serviços, produtos e marcas, alimentando ainda mais esse vasto banco de informações da World Wide Web. (GUEDES; AFONSO; MAGALHÃES, 2010).

Existem três frentes que categorizam a Mineração de Dados na Web, a Mineração de Uso, a Mineração de Estrutura e a Mineração de Conteúdo. A mineração de uso aborda a mineração das informações de uso da Web, são as informações sobre como o usuário interage com a Web. Nessa categoria são tratadas questões como personalização, interfaces adaptativas e aprendizado de perfis de usuários. A mineração de estrutura aborda a mineração das informações contidas entre os documentos da Web. Os documentos da Web se relacionam basicamente através de vínculos de hipertexto, e esses vínculos escondem informações valiosas não só sobre a topologia da Web, mas também sobre como os documentos se relacionam. A mineração de conteúdo aborda a mineração dos dados contidos dentro dos documentos da Web. A grande quantidade de formatos que os dados podem assumir (textos comuns, páginas HTML, imagens, áudio, vídeo, etc.) acaba dirigindo as técnicas de mineração a serem utilizadas. (MARINHO; GIRARDI, 2005).

Esta última se estende para a Mineração de Opinião, que é responsável por minerar e classificar opiniões.

4. Mineração de opinião

A mineração de opinião, ou Opinion Mining, também conhecida como análise de sentimentos, pode ser definida como a técnica que avalia um conteúdo subjetivo emitido em linguagem natural e descobre o sentimento que é transmitido. Geralmente associado à classificação binária entre sentimentos positivos e negativos, o termo é usado de uma forma mais abrangente para significar o tratamento computacional de opinião, sentimento e subjetividade em textos. (PANG; LEE, 2002). Com o advento da web como fonte de

informações, grande parte dos usuários tem buscado nela textos que forneçam esse tipo de informação desejada, opiniões sobre alguma entidade de interesse como um produto específico, uma empresa, um lugar, uma pessoa, dentre outros. O objetivo principal é permitir que um usuário obtenha uma visão geral sobre o que outros pensam sobre o produto, sem precisar localizar e ler cada opinião feita na web. Para atingir esse objetivo, a mineração de opinião é dividida em três grandes etapas: coleta de conteúdo, classificação e sumarização dos resultados.

4.1 Coleta de Conteúdo

Etapa na qual é feita uma busca em fontes diversas, tais como artigos em sites, comentários em mídias sociais, anúncios, documentos dentre outras. É importante a utilização de técnicas avançadas de busca, visando identificar se o conteúdo encontrado trata-se de uma opinião ou um fato. Fatos por si só devem ser descartados, porém opiniões expressas em fatos devem ser mantidas. (BECKER; TUMITAN, 2005).

4.2 Classificação

A classificação é a etapa mais importante do processo e é nela que a polaridade ou orientação da opinião é definida. Esta etapa determina se uma opinião é positiva, negativa ou neutra. Na classificação ou análise de sentimentos, são as palavras opinativas que têm a maior importância, pois, através delas, é possível determinar o sentimento expresso pelo autor. Exemplos de palavras opinativas: bom, legal, ótimo, ruim, péssimo etc. (BECKER; TUMITAN, 2005).

4.3 Sumarização dos Resultados

Etapa focada na apresentação dos resultados, que podem ser de forma textual, ou gráfica. A melhor forma de representar os resultados é a gráfica, pois facilita a visualização e entendimento dos resultados sumarizados em totais e dados estatísticos. (BECKER; TUMITAN, 2005).

5. Metodologia

A metodologia de desenvolvimento deste trabalho consistiu na definição e concretização de três etapas: obtenção do corpus de comentários, arquitetura do sistema proposto e desenvolvimento.

5.1 Corpus de Comentários

Para que o presente trabalho fosse realizado, foi necessária a construção de um corpus de comentários que serviu como entrada de dados para que o sistema realizasse a mineração da opinião. Esse corpus foi formado através da rede social Facebook, utilizando fanpages de organizações reais para coleta dos dados, tais como Sony, Samsung, Panasonic, Brastemp e Electrolux. Foram consultadas as fanpages brasileiras das marcas e puderam-se obter tanto comentários positivos quanto negativos.

A rede social Facebook foi escolhida devido a sua popularidade, alto número de usuários conectados e, principalmente, a variedade de tipos de usuários, pessoas físicas e jurídicas, característica importante para esta investigação, por promover a interação entre empresas e consumidores.

O corpus construído foi inserido em uma fanpage fictícia, que simulou uma fanpage real de uma organização. Para que o sistema analise e classifique estas opiniões, basta acessá-lo, informar o login e senha do usuário no Facebook e escolher de qual fanpage deseja baixar e classificar opiniões.

Só são consideradas as publicações feitas dentro da fanpage da organização. Se a publicação for feita por um consumidor, ou seja, outro usuário do Facebook, a própria

postagem é classificada. Caso a publicação tenha sido feita pela própria fanpage, apenas os comentários desta são classificados. Essa distinção foi necessária devido à grande quantidade de usuários que expressam suas opiniões em publicações feitas pela própria empresa.

5.2 Arquitetura do Sistema

O sistema foi dividido em três fases:

a) Primeira Fase: Responsável por acessar a fanpage da empresa no Facebook, baixar os comentários publicados e salvá-los no banco de dados MySQL. O acesso a fanpage foi feito através de login e senha da conta no Facebook. Esses comentários tiveram suas polaridades classificadas na segunda fase;

b) Segunda Fase: Responsável por acessar os comentários baixados na primeira fase e classificá-los de acordo com sua polaridade. Os comentários/opiniões podem assumir caráter positivo, negativo ou neutro;

- Opinião de caráter positivo: Opiniões de caráter positivo são aquelas em que aparecem recomendações ou elogios sobre a marca ou produto. Transmitem o sentimento de aprovação ao receptor da opinião;

- Opinião de caráter negativo: Opiniões de caráter negativo são aquelas em que aparecem reclamações ou não recomendações sobre a marca ou produto. Transmitem o sentimento de reprovação ao receptor da opinião;

- Opinião de caráter neutro: Opiniões de caráter neutro são aquelas que não transmitem sentimentos positivos ou negativos ao receptor da opinião.

c) Terceira Fase: Responsável por recuperar as sentenças classificadas na segunda fase e montar um relatório conclusivo. Esse relatório é a visão final para o usuário, que poderá tomar decisões e estabelecer planos de ação mediante ao cenário apresentado.

O funcionamento do sistema é ilustrado na Figura 1, através do diagrama de processos.

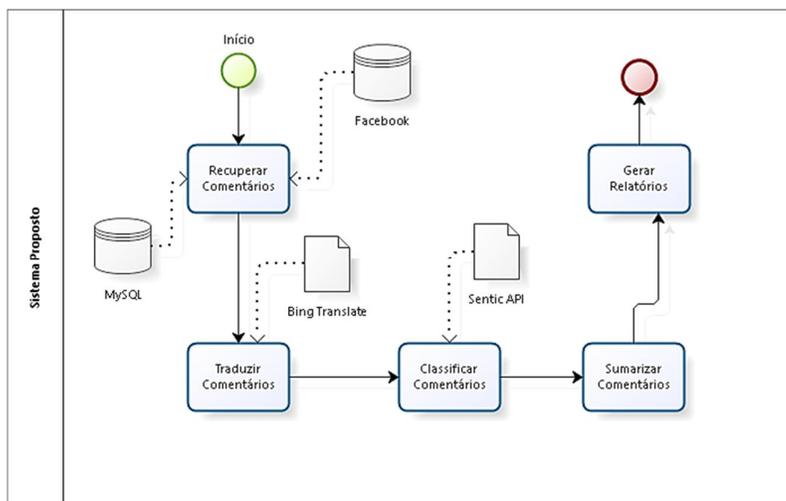


Figura 1 – Diagrama de Processos

5.3 Desenvolvimento

A primeira fase contempla a interface do sistema, a comunicação com o Facebook e o download das opiniões postadas na fanpage da empresa.

No primeiro acesso, o usuário visualizará a tela de login, onde utilizará sua conta do Facebook para logar. Em seguida, serão exibidas em uma lista, as fanpages vinculadas à conta. O usuário deverá escolher qual fanpage deseja utilizar e indicar qual será a padrão. Conforme ilustrado na Figura 2, serão então carregadas as publicações feitas na fanpage escolhida.

Bem vindo **Jéssica Silva** ▾

Organização ▾

■ Definir página atual como padrão

🏠 Posts na Fanpage

🔄 Atualizar listagem

📊 Relatórios >

Home / Posts

Listagem de Posts

	Prévia da mensagem	Usuário
Visualizar/Responder	Segunda é o dia oficial de voltar à batalha. Separa o tênis, a garrafinha de água e, de quebra, o pra dar aquela força extra no treino.	Organização
Visualizar/Responder	A menina dos olhos em qualquer ambiente!	Organização
Visualizar/Responder	A festa hoje será em casa? Então deixe a seleção de músicas por conta da sua Smart TV Samsung.	Organização
Visualizar/Responder	Não é à toa que a linha é uma das mais desejadas da Samsung.	Organização
Visualizar/Responder	Quando o assunto é oferecer ainda mais comodidade para você, a Samsung French Door Sparkling não hesita: água com gás na porta do seu refrigerador e uma dose de inovação e praticidade!	Organização
Visualizar/Responder	Fim do expediente em plena sexta-feira é aquela hora clássica de deixar a caneta cair. Mas se a caneta em questão for uma S-Pen, tem gente que não vai querer se desapegar dela por nada.	Organização
Visualizar/Responder	O HT-CT370 é uma barra de som criada especialmente para trazer mais vida à sua TV e melhorar sua experiência sonora. Fácil de instalar, ele conta com as tecnologias Bluetooth e NFC para você curtir seu som com um toque.	Organização

Figura 2 – Listagem de posts feitos na fanpage

O usuário poderá então escolher uma publicação para visualizar e responder, conforme ilustrado na Figura 3.

responder	Comprei um produto que em menos de 1 ano e meio de uso está completamente enferrujado e a pintura interna descascou inteira!	Ricardo Henrique Soares	☹️
responder	Ótimo atendimento e apoio da equipe pós venda. Recomendo!	Jéssica Rodrigues	😊
responder	Produto ruim.	Jéssica Rodrigues	☹️

Resposta

Ok. Obrigada!

Enviar Voltar a listagem

Figura 3 – Visualizando/Respondendo um comentário

Conforme mostrado nas Figuras 2 e 3, o sistema permite o gerenciamento das publicações feitas nas fanpages de uma conta da rede Facebook. A comunicação com o Facebook ocorre através da Application Programming Interface (API) disponibilizada pelo próprio Facebook à desenvolvedores de software. Como a API foi desenvolvida na linguagem de programação PHP, a mesma linguagem foi utilizada para o desenvolvimento da primeira fase deste sistema. O MySQL foi utilizado para salvar as opiniões postadas na fanpage, eliminando a necessidade de várias consultas ao Facebook, o que deixaria o processamento mais lento. Esse banco foi escolhido devido a sua facilidade de utilização e sua grande compatibilidade com as linguagens de programação utilizadas neste sistema, PHP na primeira fase e Java nas segunda e terceira fases.

Na segunda fase, cada opinião recuperada é classificada entre positiva, negativa ou neutra, de acordo com o sentimento transmitido. A Figura 4 mostra onde é exibida essa informação.

Visualizar/Responder Post

Quando o assunto é oferecer ainda mais comodidade para você, a Samsung French Door Sparkling não hesita: água com gás na porta do seu refrigerador e uma dose de inovação e praticidade! [responder]

Comentários			
	Mensagem	Usuário	Polaridade
responder	Empresa confiável e de boa reputação. Voltaria a fazer negócio.	Jéssica Rodrigues	😊
responder	PROPAGANDA ENGANOSA!!! Consultem o site RECLAME AQUI!	Jéssica Rodrigues	😡
responder	Beta melhor em tudo que faz,e parabéns pra nós .	Jéssica Rodrigues	😊
responder	Decepcionada com a Beta, as propagandas são lindas e maravilhosas, tenho 2 meses e pouco de uso do meu aparelho celular e já apresentou defeito.	Jéssica Rodrigues	😊
responder	TOTAL DESCASO COM O CLIENTE, MANDO PRA ASSISTENCIA E ME DEVOLVEM O PRODUTO ESTRAGADO E AGORA NÃO QUEREM REPARAR O ERRO!!!!!!	Jéssica Rodrigues	😡
responder	OLÁ BETA! GOSTARIA DE DEIXAR BEM CLARO AQUI PARA OUTRAS PESSOAS COMO O SEUS SERVIÇOS DE ATENDIMENTO AO CONSUMIDOR É INOPERANTE E DESQUALIFICANTE COM SEUS CLIENTES.	Jéssica Rodrigues	😊

Figura 4 – Opiniões classificadas

A classificação se dá através da API (Sentic API) do SenticNet , software criado pelo Sentic Team, grupo multidisciplinar de pesquisa dentro da NTU School of Computer Science.O objetivo era fazer com que a informação conceitual transmitida via linguagem natural fosse facilmente acessível às máquinas.

A Sentic API disponibiliza estruturas no formato XML onde armazena grande quantidade de expressões na língua inglesa. As expressões são classificadas com uma polaridade que varia de -1 a 1, onde 1 é o mais positivo possível e -1 o mais negativo possível. O zero (0) classifica a expressão como neutra. Por exemplo, a expressão "friday night", expressão que traz sentimento positivo, possui polaridade +0.728. Já a palavra "monday", dia da semana não tão popular, tem valor -0.847.

A fim de obter uma melhor performance junto a Sentic API, a sentença é dividida em partes e passada dessa forma. Cada parte recebe um score de polaridade e ao final é calculada uma média das pontuações. A média indica a polaridade da sentença como um todo.

Como a Sentic API possui apenas expressões em inglês e os comentários coletados são em português, foi necessária a utilização de uma ferramenta de tradução. Por ser gratuita, foi adotada a ferramenta Bing Translator, do sistema de buscas Bing da Microsoft. Basicamente, esta ferramenta é uma API que recebe a sentença e a língua para qual se deseja traduzir e a retorna traduzida. A eficácia deste recurso foi validada no momento do desenvolvimento desta fase do sistema, e apresentou resultado satisfatório para a tradução, não havendo perdas no sentido das sentenças.

Com as sentenças devidamente classificadas, a terceira fase do sistema se encarrega de tratar essas informações e montar uma visão conclusiva sobre o cenário da organização na rede social Facebook.

As opiniões classificadas são agrupadas e exibidas em relatórios gráficos, possibilitando a visualização da intensidade dos sentimentos. Foi desenvolvida também uma visão de opinião por consumidor, permitindo a organização saber quantas vezes cada consumidor comentou em sua página. Por fim, uma visão mensal e anual trazendo os sentimentos por período.

6. Resultados

O sistema proposto foi submetido a testes de precisão para definição de sua eficiência. A precisão é medida, considerando o número de acertos do sistema, ou seja, se o sistema classifica corretamente as opiniões, previamente rotuladas por juízes humanos, em positivas ou negativas. No treinamento, foram cadastradas na fanpage uma amostra de cem opiniões, cinquenta positivas e cinquenta negativas. A Figura 5 mostra em forma gráfica o resultado obtido.

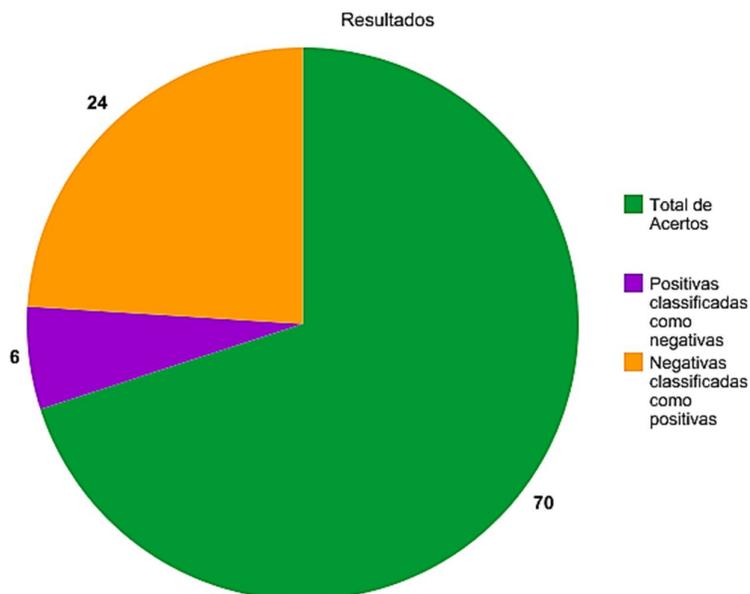


Figura 5 – Gráfico de Resultados

Foram computados um total de 70% de acertos das cem opiniões classificadas, número expressivo que mostra o bom desempenho do sistema. Da amostra, seis classificações foram classificadas erroneamente como negativas e vinte e quatro como positivas. Alguns desses resultados inesperados serão explicados a seguir.

Opinião	Juízo Humano	Sistema Proposto
Gostei muito do produto, mas a câmera frontal dele não é 100%.	Positivo	Negativo
Celular espetacular, mas falha num requisito fundamental, durabilidade.	Negativo	Positivo
É um excelente Celular, navegação internet excelente, tela nitida, câmera ótima, não trava, especialmente na hora de fazer selfs com o guia é um diferencial e tanto.	Positivo	Positivo
Tenho um fone sony há mais de um ano e sem dúvidas e muito show.	Positivo	Positivo
Sony é um lixo não compre nada, é uma ilusão.	Negativo	Negativo
Quero q vcs cubram a garantia do meu ps4 a cada dia se tornando uma empresa horrível com o trato com o cliente	Negativo	Negativo

Figura 6 – Precisão do Classificador

Na primeira opinião da Figura 6, o termo “produto” possui uma polaridade positiva de 0.154 na Sentic API. Já os termos “câmera” e “frontal” possuem respectivamente polaridades negativas de -0.026 e -0.674, valores que definem o sentimento negativo da opinião.

Por juízo humano, a polaridade desta sentença foi definida como positiva, porém, a opinião carrega a preposição adversativa “mas”, que cria um contraste de ideias na frase. Com

este contraste, podemos enxergar duas opiniões e sentimentos na opinião: antes do “mas”, “Gostei muito do produto”, com um sentimento positivo e depois do “mas”, “a câmera frontal dele não é 100%.”, com um sentimento negativo.

Na segunda opinião da Figura 6, o motivo do resultado não esperado é o mesmo da primeira opinião.

Logo, pode-se concluir que tratam-se de sentenças com um nível de dificuldade alto de identificação de um único sentimento.

As demais opiniões apresentadas, mostram casos de sucesso na classificação de acordo com juízo humano.

7. Considerações finais

O presente trabalho teve como objetivo a pesquisa e implementação da mineração de opinião na rede social Facebook. A implementação se deu através de um software que obteve um resultado satisfatório diante do esperado, sendo capaz de classificar corretamente 70% das opiniões da amostra apresentada. Apesar das opiniões do corpus montado serem reais, novos testes podem ser realizados, utilizando uma fanpage de uma organização real, com uma marca expressiva de comentários. Caso os resultados continuem satisfatórios nesse cenário, a ferramenta poderia começar a ser utilizada por empresas.

Porém, mesmo com um bom resultado, existem alguns pontos que podem ser melhorados, como, por exemplo, a velocidade do processamento da classificação. Atualmente, esse processamento é um tanto oneroso em questão de tempo, e esse fator é hoje muito escasso para as empresas. Outro ponto a ser melhorado é a utilização da classificação através de janelas, o que corresponderia ao envio de mais de uma palavra (formando expressões) para obtenção da polaridade na Sentic API. Atualmente, o sistema proposto envia uma palavra por vez. A Sentic API oferece suporte para esta melhoria, possuindo expressões como “a lot of books”, que significa “muitos livros” e tem polaridade positiva de 0.047. Outro exemplo é “a lot of noise” que significa “muito barulho”, com polaridade negativa de -0.619. Essa implementação provavelmente traria uma classificação ainda mais precisa do sentimento, pois carrega um conhecimento mais específico.

Apesar das limitações citadas, o foco particular do processamento da língua portuguesa, está entre as principais contribuições deste trabalho, não só para a validação de um método e criação de um sistema de análise de sentimentos, mas principalmente porque há poucos recursos para esta língua, uma vez que a maior parte das pesquisas envolve o idioma inglês.

Desta forma, conclui-se que o presente trabalho contribuiu com conhecimento para o segmento de Análise de Sentimentos da Inteligência Artificial e Ciência da Computação, explorando mais uma forma de implementar a interpretação de linguagem natural e identificação de sentimentos através de um software. Além disso, apesar de ainda precisar de melhorias, o software pode ser utilizado e aliado das organizações, provendo informações sobre as marcas, através da classificação das opiniões dos consumidores.

Referências

- BECKER, K.; TUMITAN, D. Introdução à Mineração de Opiniões: Conceitos, Aplicações e Desafios. UFRGS, 2005. Disponível em: <http://www.inf.ufrgs.br/~kbecker/lib/exe/fetch.php?media=minicursosbbd_versaosubmetida.pdf>. Acesso em: 10 maio 2015.
- COOLEY, R. Web mining: information and pattern Discovery on the World Wide Web, Proceedings of the 9th IEEE International Conference on Tools with Artificial Intelligence. IEEEXplore, 1997. Disponível em:

- <<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=632228>>. Acesso em: 10 maio 2015.
- GUEDES, R.; AFONSO, D.; MAGALHÃES, L. H. de. Mineração de opiniões de usuários na busca de conhecimento. *Vianna Sapiens*, 2010. Disponível em: <http://www.viannajunior.edu.br/files/uploads/20131001_141137.pdf>. Acesso em: 4 maio 2015.
- LEVY, P. A inteligência coletiva: por uma antropologia do ciberespaço. São Paulo: Loyola, 2000.
- MARINHO, L. B.; GIRARDI, R. Mineração na Web. *Research Gate*, 2005. Disponível em: <http://www.researchgate.net/profile/Leandro_Marinho/publication/228369738_Minerao_na_Web/links/0deec539852e70896c000000.pdf>. Acesso em: 10 maio 2015.
- OLIVEIRA, A. C. Inteligência competitiva na internet. Rio de Janeiro: Brasport, 2006.
- PANG, B.; LEE, L.; VAITHYANATHAN, S. Thumbs up? Sentiment Classification using Machine Learning Techniques. *ACM DIGITAL LIBRARY*, 2002. Disponível em: <<http://dl.acm.org/citation.cfm?id=1118704>>. Acesso em: 04 jun. 2015.
- PORTER, M. Estratégia competitiva. Rio de Janeiro: Campus, 1986.
- ROSA, J. L. G. Fundamentos da Inteligência Artificial. Rio de Janeiro: Gen-LTC, 2011.
- SILVA, B. C. D. da. et al. Introdução ao Processamento das Línguas Naturais e Algumas Aplicações. 2007. 119 f. Série de Relatórios do Núcleo Interinstitucional de Linguística Computacional, São Carlos, 2007. Disponível em: <<http://www.lettras.etc.br/ebralc/NILCTR0710-DiasDaSilvaEtAl.pdf>>. Acesso em: 4 maio 2015.