

UNIVERSIDADE FEDERAL DO PARANÁ

ANTONIO JOSÉ HABLE FOIT

**DESCRIÇÃO DO PROCESSO DE ANÁLISE DE OPINIÕES NO TWITTER:  
BITCOIN NO CENÁRIO BRASILEIRO**

CURITIBA

2017

ANTONIO JOSÉ HABLE FOIT

**DESCRIÇÃO DO PROCESSO DE ANÁLISE DE OPINIÕES NO TWITTER:  
BITCOIN NO CENÁRIO BRASILEIRO**

Trabalho de conclusão de curso apresentado como requisito parcial à obtenção do grau de Bacharel em Gestão da Informação no curso de graduação em Gestão da Informação, Setor de Ciências Sociais Aplicadas da Universidade Federal do Paraná.

Orientador: Profº. Drº. José Simão de Paula Pinto.

CURITIBA

2017

## **TERMO DE APROVAÇÃO**

ANTONIO JOSÉ HABLE FOIT

### **DESCRIÇÃO DO PROCESSO DE ANÁLISE DE OPINIÕES NO TWITTER: BITCOIN NO CENÁRIO BRASILEIRO**

Trabalho apresentado como requisito parcial à obtenção do grau de bacharel em Gestão da Informação no curso de graduação em Gestão da informação, Setor de Ciências Sociais Aplicadas, Universidade Federal do Paraná, pela seguinte banca examinadora:

---

Prof.º Dr.º José Simão de Paula Pinto

Orientador - Setor de Ciências Sociais Aplicadas da Universidade  
Federal, UFPR

---

Prof. Me. André José Ribeiro Guimarães

Setor de Ciências Sociais Aplicadas da Universidade Federal, UFPR

---

Prof. Dr.º Egon Walter Wildauer

Setor de Ciências Sociais Aplicadas da Universidade Federal, UFPR

Curitiba, 04 de dezembro de 2017

## **AGRADECIMENTOS**

Agradeço primeiramente aos meus pais, que sempre me apoiaram e me ajudaram em todos os momentos que precisei. Aos esforços que desempenharam para garantir que eu sempre tivesse uma boa educação e que nunca desistisse dos meus sonhos.

À minha namorada, que esteve comigo durante os quatro anos do curso e que sempre me incentivou nas horas mais difíceis, sempre demonstrando amor e carinho.

Quero agradecer de modo geral a todos os professores do curso, que de uma maneira ou outra, contribuíram no meu desenvolvimento acadêmico e profissional. De maneira especial, gostaria de agradecer ao professor Simão, que durante a realização deste trabalho, demonstrou ser um grande orientador, apresentando bons conselhos e soluções para que fosse possível a finalização do mesmo.

Obrigado também a todos os amigos que fiz na faculdade, os quais estiveram comigo durante longo quatro anos, em especial Lucas Pol, Ruan Torres, Gabriel Monteiro, André Zani, Alan Falcoski, Ivan Cardoso, Murilo Kenji e Paulo Moreira. Com vocês tudo tornou-se mais divertido e espero manter esta amizade para sempre.

De modo geral gostaria de agradecer a todos os demais amigos, os quais trago comigo desde os tempos de escola, sobretudo a Joice Theodoroski, a qual sempre me apoiou nos momentos necessários e me ajudou para que fosse possível terminar este trabalho. Também a Wellisson Ribeiro, o qual me ajudou e muito na elaboração do script para coleta dos dados do Twitter, sem o apoio deles, a finalização deste trabalho não seria possível.

Também gostaria de agradecer à minha equipe de trabalho, em especial a Samuel Nunes, o qual como supervisor, demonstrou ser um grande amigo, me ensinando sobre diversas coisas e me aconselhando de maneira sem igual.

Enfim agradeço a todos aqueles que estiverem comigo durante estes quatro anos. Obrigado de coração pelo apoio, torcida e ajuda.

## RESUMO

Desde os primórdios, as pessoas sentem a necessidade de conhecer a opinião de outras pessoas sobre alguns assunto, a fim de melhorar as chances de uma assertiva tomada de decisão. Nos dias atuais, estas opiniões estão cada vez mais acessíveis, de modo que é possível encontrá-las em publicações em diversas redes sociais. As redes sociais possibilitam que as pessoas se identifiquem com grupos e se sintam livres para compartilhar opiniões sobre produtos, serviços, empresas ou até mesmo uma tendência, como é o caso do Bitcoin. No entanto, por serem muito difundidas e acessadas, acabam gerando uma grande quantidade de registros, inviabilizando o entendimento manual desses dados, sendo necessárias técnicas automatizadas para realizar tal processo. O nome atribuído para tal processo é análise de sentimento ou mineração de opinião, o qual é utilizado para classificar as opiniões em positivas, negativas ou neutras. Uma vez classificadas, é possível entender de maneira geral qual o sentimento sobre tal assunto e este tipo de análise pode ser fundamental no processo de decisão. Este trabalho apresenta uma maneira simples e acessível para realizar a mineração de opiniões no Twitter, sendo mostrado as etapas de coleta e tratamento dos dados até a aplicação da ferramenta Semantria para a análise dos mesmos. O principal objetivo se concentra em encontrar as dificuldades e facilidades na realização deste processo. Uma das dificuldades encontradas é a capacidade da ferramenta em tratar gírias, ironias e abreviações, os quais são muito comuns em redes sociais. Como facilidades podemos observar a disponibilidades de APIs para a coleta de informações das redes sociais e uma ampla quantidade de tutoriais que exemplificam como utilizar a ferramenta escolhida neste projeto.

Palavras-Chaves: Mineração de Opiniões. Análise de sentimento. Bitcoin. Criptomoedas. Redes sociais. Twitter.

## **ABSTRACT**

Since the earliest times the people felt the need to know opinions of other people about some subject, in order to improve the chances of an assertive decision-making. Nowadays, these opinions are becoming increasingly accessible, once it is possible to find them on several posts in the social networks. Social networks allow people to identify with groups and feel free to share opinions about products, services, companies or even a trend, such as Bitcoin. However, because they are very widespread and accessed, they end up generating a large amount of records, making it impossible to manually understand these data, and automated techniques are required to carry out this process. The name assigned for this process is sentiment analysis or opinion mining, which is used to rank opinions as positive, negative, or neutral. Once classified, it is possible to understand in general, the sentiment about such subject and this kind of analysis can be fundamental in the decision process. This paper presents a simple and accessible way to perform the mining of opinions on Twitter, showing the steps to collect and process data until the application of the Semantria tool to analyze them. The main objective is to find the difficulties and facilities in such process. One of the difficulties found is the ability of the tool to handle slang, ironies and abbreviations, which are very common in social networks. As facilities we can observe the availability of APIs to collect information from social networks and large amount of tutorials that exemplify how to use the tool chosen in this project.

Keywords: Opinion Mining. Sentiment Analysis. Bitcoin. Cryptocurrency. Social Network. Twitter.

## LISTA DE FIGURAS

FIGURA 1	–	RESULTADO DA PESQUISA REALIZADA NO SITE SOCIALMENTION .....	13
FIGURA 2	–	TERMOS UTILIZADOS EM PESQUISA NA WEB OF SCIENCE E SCOPUS E RESPECTIVO RESULTADO.....	14
FIGURA 3	–	ÁREAS DE ESTUDO COM MAIORES NÚMEROS DE TRABALHOS RELACIONADOS AO TERMO MINERAÇÃO DE OPINIÕES NA BASE WEB OF SCIENCE.....	14
FIGURA 4	–	ÁREAS DE ESTUDO COM OS MAIORES NÚMEROS DE TRABALHOS RELACIONADOS AO TERMO MINERAÇÃO DE OPINIÕES NA BASE SCOPUS .....	15
FIGURA 5	–	PÁGINA DO PERFIL DE UM DOS CRIADORES DO TWITTER.....	23
FIGURA 6	–	BLOCO 0 - THE GENESIS BLOCK .....	27
FIGURA 7	–	MAPA DE CALOR DA ACEITAÇÃO MUNDIAL DO BITCOIN .....	32
FIGURA 8	–	PROCESSO DE COLETA, TRATAMENTO E ANÁLISE DE SENTIMENTO EXPRESSOS EM TWEETS. ....	34
FIGURA 9	–	INFORMAÇÕES NECESSÁRIAS PARA CRIAR UMA APLICAÇÃO... ..	37
FIGURA 10	–	CHAVES E TOKENS DA APLICAÇÃO.....	38
FIGURA 11	–	RETORNO DE UMA CONSULTA REALIZADA COM O SCRIPT CRIADO .....	41
FIGURA 12	–	EXEMPLO DE DEFINIÇÃO DE CODIFICAÇÃO UTF-8 NO HTML .	42
FIGURA 13	–	ABA LEXALYTICS RESPONSÁVEL PELAS OPÇÕES DE MINERAÇÃO DE SENTIMENTOS DO PLUGIN.....	44
FIGURA 14	–	PAINEL DE OPÇÕES DO PLUGIN SEMANTRIA .....	45
FIGURA 15	–	ANÁLISE REALIZADA COM O PLUGIN SEMANTRIA.....	46
FIGURA 16	–	CLASSIFICAÇÃO DAS POLARIDADES DOS TWEETS ANÁLISADOS .....	49
FIGURA 17	–	CÓDIGO PHP UTILIZADO PARA RESTRINGIR ATRIBUTOS .....	53
FIGURA 18	–	EXEMPLO DE IRONIA CLASSIFICADA INCORRETAMENTE .....	57
FIGURA 19	–	EXEMPLO DE EQUÍVOCO EM RELAÇÃO A PALAVRAS DE BAIXO CALÃO.....	57

## LISTA DE QUADROS

QUADRO 1 – PARAMETROS UTILIZADOS DO MÉTODO GET EARCH/TWEETS . .....	39
QUADRO 2 – METADADOS COLETADOS QUE REFERENCIAM A PUBLICAÇÃO .....	40
QUADRO 3 – DIFICULDADES E FACILIDADES NA ETAPA DO PRÉ-PROJETO	50
QUADRO 4 – DIFICULDADES E FACILIDADES NA ETAPA DE COLETA E ARMAZENAMENTO DOS DADOS .....	51
QUADRO 5 – DIFICULDADES E FACILIDADES NA ETAPA DE TRATAMENTO DOS DADOS.....	54
QUADRO 6 – DIFICULDADES E FACILIDADES NA ETAPA DE MINERAÇÃO DE OPINIÕES.....	56
QUADRO 7 – COMPARAÇÃO ENTRE ANÁLISE MANUAL E ANÁLISE AUTOMÁTICA DE TWEETS.....	58

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	10
1.1	PROBLEMATIZAÇÃO	10
1.2	OBJETIVOS	12
1.2.1	Objetivo Geral	12
1.2.2	Objetivos Específicos	12
1.3	JUSTIFICATIVA	12
1.4	DELIMITAÇÃO DA PESQUISA	15
1.5	ESTRUTURA DO DOCUMENTO	16
<b>2</b>	<b>REFERENCIAL TEÓRICO</b>	17
2.1	MINERAÇÃO DE OPINIÃO	17
2.1.1	Opinião	17
2.1.2	Mineração de Opinião: Conceituação	18
2.1.3	Etapas da Mineração de Opinião	19
2.1.3.1	Subjetividade	19
2.1.3.2	Extração	20
2.1.3.3	Classificação	20
2.1.3.4	Sumarização	21
2.2	REDES SOCIAIS	21
2.2.1	Twitter	22
2.2.2	Facebook	24
2.3	CRIPTOMOEDAS	24
2.4	BITCOIN	25
2.4.1	O Que É Bitcoin	25
2.4.2	Como surgem os Bitcoins	27
2.4.3	Transações De Bitcoins	28
2.4.4	Características Do Bitcoin	30
2.4.5	Bitcoin no Brasil Hoje	31
<b>3</b>	<b>PROCEDIMENTOS METODOLÓGICOS</b>	34
3.1	DEFINIÇÃO DA POPULAÇÃO	34
3.2	COLETA E ARMAZENAMENTO DOS DADOS	35
3.2.1	API REST do Twitter	35

3.2.2	Criando a aplicação .....	36
3.2.3	Executando a aplicação .....	38
3.3	TRATAMENTO DOS DADOS .....	42
3.4	MINERAÇÃO DE OPINIÃO .....	43
3.4.1	Instalação do Plugin Semantria para Excel .....	44
3.4.2	Etapas da mineração de opiniões com o Plugin Semantria para Excel.....	44
3.5	FERRAMENTAS NECESSÁRIAS .....	46
<b>4</b>	<b>APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS.....</b>	<b>48</b>
4.1	ANÁLISE DA ETAPA DE SUMARIZAÇÃO.....	48
4.2	FACILIDADES E DIFICULDADES ENCONTRADAS NA ANÁLISE DE SENTIMENTO .....	50
<b>5</b>	<b>CONSIDERAÇÕES FINAIS .....</b>	<b>59</b>
5.1	TRABALHOS FUTUROS.....	60
	<b>REFERÊNCIAS .....</b>	<b>62</b>
	<b>APÊNDICE A – RETORNO DE UMA CONSULTA NO FORMATO JSON</b>	<b>66</b>
	<b>APÊNDICE B – SCRIPT PARA COLETA DE TWEETS.....</b>	<b>70</b>

## 1 INTRODUÇÃO

As mudanças tecnológicas que vêm ocorrendo no mundo atualmente, permitiram o surgimento de tendências de ordem mundial, as quais, revolucionaram a maneira como realizamos nossas atividades. Um desses casos é o da criptomoeda Bitcoin, a qual nas palavras de Ulrich (2014, p. 15), “é uma forma de dinheiro, assim como o real, o dólar ou o euro, com a diferença de ser puramente digital e não ser emitido por nenhum governo”.

Desta maneira, o Bitcoin vem ganhando força em nossa sociedade, devido a diversas características que o distingue das demais moedas, como por exemplo, as pequenas taxas de transações e a velocidade de execução das mesmas. O anonimato também é um fator chamativo, no entanto, a forma como é administrada, em que não é necessário o envolvimento de nenhum terceiro no processo, é a sua principal força e motivo de criação.

Atualmente a moeda Bitcoin tem ganho muita força na internet, onde sites especializados estão postando cada vez mais notícias e novidades sobre o tema. Porém, os grandes meios de difusão deste conceito são as redes sociais, que se mostram propagadores de conteúdos em nossa sociedade. Além disso, pessoas frequentemente expressam suas opiniões ou procuram por opiniões de conhecidos e especialistas nestas redes, o que justifica estudos na área de análise de sentimentos.

Para Bertaglia (2015, p. 1), “A análise de sentimentos surge como uma alternativa para lidar com a grande quantia de opiniões disponíveis, visando facilitar a compreensão das informações contidas nelas”. Seu principal objetivo é identificar automaticamente opiniões expressas em textos e os sentimentos relacionados a elas.

Tendo em vista as aplicações da análise de sentimentos – também conhecida como mineração de opinião – este trabalho objetiva apresentar um roteiro para a realização de uma pesquisa baseada em tweets, mostrando as facilidades e dificuldades do processo. Para tanto, foi escolhido como tema de pesquisa a criptomoeda Bitcoin.

### 1.1 PROBLEMATIZAÇÃO

O Bitcoin é uma moeda virtual que dispensa a participação de terceiros na realização de atividades monetárias, de modo que os únicos participantes do processo

são: o indivíduo que possui a moeda e o indivíduo que irá receber a mesma. O Bitcoin pode ser considerado um ativo financeiro, pois é utilizado como moeda de troca em diversos países que já perceberam a sua importância. Além disso, o Bitcoin pode ser usado como investimento de risco, devido o valor ser definido de acordo com a oferta e demanda do mercado.

De acordo com o site Bitcoin.org (2017), grande parte da confiança no Bitcoin vem do fato de não exigir nenhuma confiança em tudo, pois é totalmente open-source e descentralizado, permitindo que qualquer pessoa tenha acesso ao código-fonte completo, podendo verificar exatamente como funciona a moeda. Além do mais, nenhuma organização ou indivíduo pode controlar o Bitcoin, fazendo com que a rede continue segura, mesmo que nem todos os seus usuários possam ser confiáveis.

Os usuários que utilizam esta moeda – por ser gerenciada em meio virtual – costumam expressar suas opiniões em postagens nas redes sociais na internet, trazendo muitas abordagens e assuntos diferenciados.

As opiniões de um indivíduo ou grupo de pessoas podem acabar influenciando a maneira como outras pessoas pensam ou agem, apresentando um motivo para buscar opiniões que auxiliem na tomada de decisão. Com o advento das redes sociais, a exposição e a busca de opiniões tornou-se algo rotineiro, devido a sua velocidade, diversidade de usuários e abrangência mundial. Assim, usuários não estão mais restritos à opinião de um amigo ou conhecido a respeito de algo, na medida em que há uma grande quantidade de informação publicamente disponível. (LIU, 2012).

Desse modo, realizar um monitoramento das redes sociais e a partir delas extrair informações relevantes, tornou-se um grande desafio, uma vez que analisar os textos manualmente é uma tarefa quase impossível devido a amplitude de postagens existentes, além de que, os registros encontrados na web estão na forma de textos não estruturados. (LIU, 2012).

Estas postagens podem ser recuperadas através da utilização de ferramentas que as próprias redes sociais possuem. Feito esta etapa, algumas técnicas de mineração de opinião são empregadas para determinar a polaridade que categoriza a opinião em positiva, negativa ou neutra, facilitando o entendimento da compreensão de um determinado grupo sobre um assunto.

A partir deste relato, o presente trabalho pretende resolver o principal problema deste estudo: Quais as principais dificuldades e facilidades encontradas na realização de uma análise de sentimentos na rede social Twitter?

## 1.2 OBJETIVOS

De acordo com o problema de pesquisa, foi possível definir os objetivos deste trabalho, os quais estão divididos em um objetivo geral e três objetivos específicos.

### 1.2.1 Objetivo Geral

Levantar as dificuldades e facilidades da realização de uma análise de sentimentos no Twitter utilizando como tema de pesquisa o Bitcoin.

### 1.2.2 Objetivos Específicos

Os objetivos específicos definidos para alcançar o objetivo geral deste estudo são os seguintes:

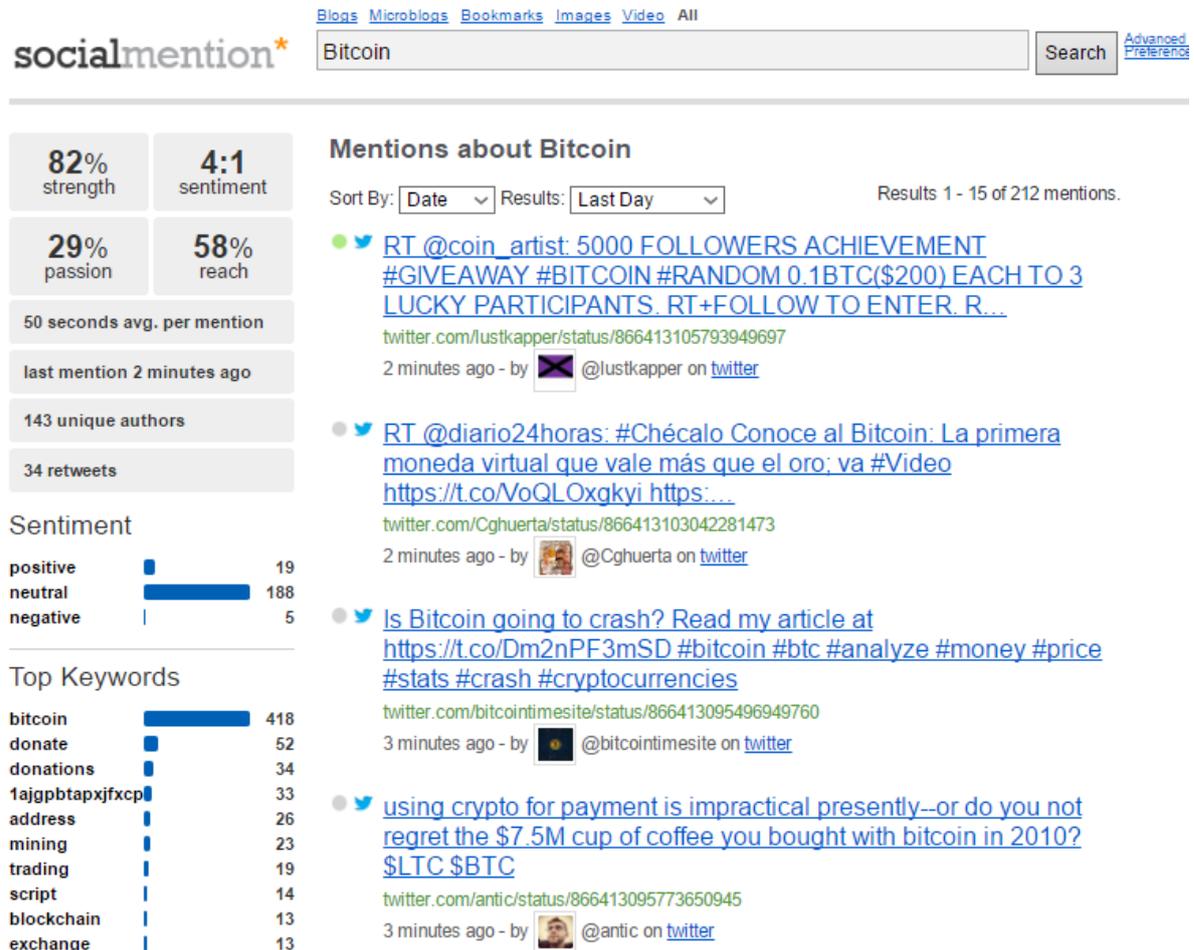
- a) Levantar referencial teórico sobre Bitcoin para pesquisa teste realizada no Twitter;
- b) Criar um roteiro com os procedimentos necessários para realizar uma análise de sentimento no Twitter;
- c) Elaborar quadros com as dificuldades e facilidades encontradas pelo autor com a realização da pesquisa.

## 1.3 JUSTIFICATIVA

Conforme pesquisa com o termo Bitcoin realizada na ferramenta SocialMention (2017), verificou-se que a força de probabilidade de pessoas discutirem o tema nas redes sociais é de 82%. No momento da pesquisa, foram retornadas 212 menções ao termo, das quais 19 possuíam sentimento positivo, 5 negativas e 188 neutras, o que pode significar que as pessoas não se posicionaram sobre o tema ainda, o que possibilita a realização de trabalhos sobre o assunto, para entender o

posicionamento dos usuários de redes sociais em relação ao Bitcoin. O resultado da Pesquisa pode ser visto na FIGURA 1.

FIGURA 1 - RESULTADO DA PESQUISA REALIZADA NO SITE SOCIALMENTION



FONTE: O Autor (2017).

Para justificar o trabalho em uma abordagem científica, foi realizada uma pesquisa em março de 2017 nas bases Web of Science – Coleção Principal (Thomson Reuters Scientific) e na Scopus (Elsevier) com a intenção de verificar trabalhos relacionados com a mineração de opinião sobre a moeda Bitcoin. A FIGURA 2 apresenta os termos utilizados na pesquisa e os resultados obtidos em cada base. É possível perceber que nas duas bases não foi recuperado nenhum resultado dos termos “Opinion mining” e “Bitcoin” além de um pequeno número de publicações cruzando os termos “Opinion”, “Sentiment” e “Sentiment Analysis” com o termo “Bitcoin”, o que representa uma falta de pesquisas nesta área de estudo, salientando a importância da realização deste trabalho.

FIGURA 2 - TERMOS UTILIZADOS EM PESQUISA NA WEB OF SCIENCE E SCOPUS E RESPECTIVO RESULTADO

Termos	Resultados obtidos em cada Base	
	Scopus	Web of Science
"Opinion Mining"	8,027	1,221
"Sentiment Analysis"	12,952	2,913
"Bitcoin"	1,436	369
"Opinion" AND "Bitcoin"	7	2
"Sentiment" AND "Bitcoin"	8	2
"Sentiment Analysis" AND "Bitcoin"	3	0
"Bitcoin AND Opinion Mining"	0	0

FONTE: O Autor (2017).

A FIGURA 3 mostra as principais áreas de estudo da base Web of Science onde se desenvolveram pesquisas em mineração de opiniões.

FIGURA 3 - ÁREAS DE ESTUDO COM MAIORES NÚMEROS DE TRABALHOS RELACIONADOS AO TERMO MINERAÇÃO DE OPINIÕES NA BASE WEB OF SCIENCE

COMPUTER SCIENCE INFORMATION SYSTEMS (466)
COMPUTER SCIENCE ARTIFICIAL INTELLIGENCE (450)
COMPUTER SCIENCE THEORY METHODS (410)
ENGINEERING ELECTRICAL ELECTRONIC (336)
COMPUTER SCIENCE INTERDISCIPLINARY APPLICATIONS (120)
TELECOMMUNICATIONS (101)
COMPUTER SCIENCE SOFTWARE ENGINEERING (91)
OPERATIONS RESEARCH MANAGEMENT SCIENCE (58)
INFORMATION SCIENCE LIBRARY SCIENCE (57)
COMPUTER SCIENCE HARDWARE ARCHITECTURE (57)
LANGUAGE LINGUISTICS (55)
LINGUISTICS (46)
MANAGEMENT (32)
BUSINESS (30)

FONTE: O Autor (2017).

A FIGURA 4 mostra as principais áreas de estudo da Scopus onde se desenvolveram pesquisas em mineração de opiniões.

FIGURA 4 - ÁREAS DE ESTUDO COM OS MAIORES NÚMEROS DE TRABALHOS RELACIONADOS AO TERMO MINERAÇÃO DE OPINIÕES NA BASE SCOPUS

Computer Science	(1,856) >
Engineering	(424) >
Mathematics	(405) >
Social Sciences	(226) >
Decision Sciences	(193) >
Business, Management and Accounting	(151) >
Arts and Humanities	(115) >
Medicine	(34) >
Physics and Astronomy	(25) >
Multidisciplinary	(24) >

FONTE: O Autor (2017).

Dessa maneira, é possível verificar que foram encontrados poucos trabalhos relevantes envolvendo as duas áreas – Bitcoin e análise de sentimentos – justificando mais uma vez a realização do estudo.

Por fim, em uma abordagem pessoal do autor, a possibilidade de aprender sobre a área de mineração de opinião e também sobre o Bitcoin foi fator determinante para a motivação deste trabalho, principalmente pelo valor da moeda estar batendo os 43 mil reais no momento em que o trabalho é escrito. Dessa maneira, o mesmo acredita que a experiência obtida com a pesquisa sobre a criptomoeda pode auxiliar em investimentos e as etapas que fazem parte da mineração de opinião podem ser utilizadas futuramente em outros projetos.

#### 1.4 DELIMITAÇÃO DA PESQUISA

Esta pesquisa possui limitações quanto à sua coleta de dados, a qual foi realizada somente na rede social Twitter, focando em tweets do ano de 2017 para a definição da base de dados a ser trabalhada. Além disso, a linguagem dos tweets coletados foi somente em português.

A mineração de opinião em sua totalidade – tipos, metodologias, algoritmos e ferramentas – não foi abordada, restringindo somente a explicação dos assuntos necessários para alcançar o objetivo deste estudo.

O mesmo ocorre para o assunto Bitcoin, em que o foco deste trabalho é apresentar informações gerais sobre esta moeda, sem adentrar em informações técnicas relacionadas à tecnologia, criptografia, algoritmo, entre tantas outras abordagens que poderiam ser estudadas.

## 1.5 ESTRUTURA DO DOCUMENTO

Este estudo está organizado nas seguintes seções:

O referencial teórico que norteia esta pesquisa é apresentado na seção 2 e trabalha definições de opiniões e a técnica de análise de sentimentos, além das redes sociais e da criptomoeda Bitcoin.

A seção 3 apresenta os procedimentos metodológicos escolhidos para alcançar os objetivos previamente estabelecidos, tais como metodologia, abordagens e ferramentas.

Na seção 4 são mostrados e discutidos os resultados e por fim, na seção 5 são apresentadas as considerações finais, juntamente com indicações para trabalhos futuros.

## 2 REFERENCIAL TEÓRICO

Nas próximas seções são apresentados conceitos que baseiam a pesquisa e auxiliaram no desenvolvimento da metodologia deste trabalho. As definições abordadas são: A mineração de opinião, redes sociais, as criptomoedas e o Bitcoin.

### 2.1 MINERAÇÃO DE OPINIÃO

Nos tópicos seguintes, serão apresentados conceitos que dizem respeito a mineração de opiniões, uma área de estudo relativamente nova, mas que possui diversas aplicações, principalmente em redes sociais.

#### 2.1.1 Opinião

Para entender a mineração de opiniões, é importante compreender o seu objeto de estudo.

Becker e Tumitan (2013), discorrem que:

Opiniões têm grande influência sobre o comportamento das pessoas. Decisões simples como qual carro comprar, qual filme ver, ou em qual ação investir são frequentemente baseadas em opiniões de pessoas próximas, de especialistas, ou de estudos conduzidos por instituições especializadas. Organizações baseiam suas estratégias de negócio e investimentos na opinião de seus clientes sobre seus produtos ou serviços. A importância da opinião é tão grande que muitas empresas (e.g. marketing, relações públicas, pesquisas) têm seu negócio voltado à obtenção deste tipo de informação. (BECKER; TUMITAN, 2013, p. 2).

Liu (2012, p. 8, tradução nossa) complementa que “As opiniões são centrais para quase todas as atividades humanas, porque são chaves influenciadoras de nossos comportamentos. Sempre que temos a necessidade de tomar uma decisão, queremos saber a opinião dos outros.”

Becker e Tumitan (2013), complementam que

A explosão das mídias sociais alterou este cenário, disponibilizando a indivíduos e organizações conteúdo de opinião diversificado e em grandes volumes. Usuários da web têm a oportunidade de registrar e divulgar suas ideias e opiniões através de comentários, fóruns de discussão, blogs, Twitter, redes sociais, entre outros. Isto aumenta as opções dos indivíduos na busca

de opiniões, pois não estão mais limitados a sua rede pessoal de contatos (e.g. familiares, amigos, conexões profissionais) ou a opiniões de especialistas disponíveis publicamente (e.g. revistas, jornais). (BECKER; TUMITAN, 2013, p. 4).

Ainda segundo Liu (2012, p. 17, tradução nossa), “Uma opinião consiste em dois elementos chave: um alvo e um sentimento sobre este alvo.” Na mesma linha de pensamento, Ferreira (2010), diz que opiniões podem ser expressas sobre qualquer coisa (alvo), como produtos, serviços, indivíduos, eventos, etc. Comumente usa-se o termo “objeto” para referir-se ao alvo da opinião, podendo este conter diversas características ou subpartes de maneira que essas podem também ser alvo de opiniões.

De acordo com Silva (2010, p.7) “Uma opinião é uma visão, atitude, sentimento, emoção ou avaliação sobre um aspecto ou objeto por parte de um detentor de opinião, podendo ela ser positiva ou negativa”.

Becker e Tumitan (2013), afirmam que:

Os termos sentimento e opinião frequentemente são usados como sinônimo neste contexto. A polaridade de um sentimento pode ser classificada em classes discretas (e.g. positiva, negativa ou neutra), ou como um intervalo que representa a intensidade deste sentimento, tipicamente  $[-1, 1]$ . Já o termo emoção é usado para designar as percepções e pensamentos subjetivos de uma pessoa, tais como raiva, desgosto, medo, alegria, tristeza e surpresa, não representando necessariamente um posicionamento ou uma atitude em relação ao alvo. (BECKER; TUMITAN, 2013, p. 4).

Opiniões podem ser regulares – o autor da mesma expressa um sentimento, atitude, emoção ou percepção sobre um alvo (“Usar Bitcoin é muito bom”) – ou comparativas – expressam o sentimento com base na relação de similaridades ou diferenças entre duas ou mais entidades, ou preferência quanto a algum aspecto compartilhado (“O Bitcoin é muito melhor que o Litecoin”). Opiniões também podem ser diretas (“a transação é muito rápida”) ou indiretas (“minha venda está demorando depois que mudou o site”). (BECKER; TUMITAN, 2013).

### 2.1.2 Mineração de Opinião: Conceituação

Análise de sentimentos, também conhecida como mineração de opinião, é uma área de pesquisa que estuda opiniões, sentimentos, avaliações, atitudes e

emoções de pessoas em relação a determinadas entidades – como produtos, serviços, organizações, eventos e assuntos, bem como seus atributos constituintes. (LIU, 2012).

De acordo com Bertaglia (2015), existem diferentes terminologias encontradas na literatura, como mineração de opinião, análise de sentimentos, extração de opinião, mineração de sentimentos, análise de subjetividade, entre outros. No entanto, todos esses termos estão contidos na área de mineração de opinião e análise de sentimentos.

A Mineração de Opinião e Análise de Sentimento, resumidamente, analisa as opiniões, sentimentos, avaliações, apreciação descritas por pessoas referente a problemas, eventos, tópicos e seus atributos. (PANG; LEE, 2008).

Os estudos referentes à área de Mineração de Opinião são recentes, tendo seu marco inicial a partir de 2001. Surgiram como uma nova área do conhecimento a partir de reunião de Recuperação de Informação e Linguística Computacional no ambiente online. (PANG; LEE; VAITHYANATHAN, 2002).

Silva (2010), discorre que mineração de textos é uma área que tem como principal objetivo extrair conhecimento implícito de grandes quantidades de textos escritos em linguagem natural. Ainda de acordo com Silva (2010), com o aumento de pessoas que têm acesso à internet e das ferramentas de expressão de opinião, os textos com as opiniões e experiência de usuários vêm se tornando a maior parte do conteúdo da Web, transformando as ferramentas de análise de sentimento em um fator importante para se extrair informação desse emaranhado de textos.

### 2.1.3 Etapas da Mineração de Opinião

De acordo com Berlim (2015, p. 40), “minerar opiniões não se trata de uma tarefa fácil, muito pelo contrário, o processo é composto por complexas etapas, cada uma com sua importância e limitação”. Estas etapas são denominadas Subjetividade, Extração, Classificação e Sumarização e serão explicadas nos tópicos abaixo.

#### 2.1.3.1 Subjetividade

Esta etapa é iniciada com a verificação das sentenças, a fim de identificar se tais são objetivas ou subjetivas. A Objetividade está relacionado a informações

factuais, enquanto Subjetividade está relacionado a opiniões ou crenças do autor. Na identificação da opinião e detecção de subjetividade, as sentenças objetivas e subjetivas são separadas, de modo que somente as subjetivas serão utilizadas, pois elas apresentam um sentimento, emoção ou pensamento relativo a algo, ou seja, representam as opiniões que serão analisadas. (SILVA, 2010).

Berlim (2015) discorre que:

A fase subjetividade é de extrema importância para se filtrar apenas as sentenças que se configuram como opiniões e que são realmente pertinentes para o estudo nessa área, ao se utilizar uma sentença objetiva, perde-se o sentido do estudo, por isso tem-se a necessidade dessa diferenciação no início do processo, evitando-se assim desperdícios. (BERLIM, 2015, p. 41).

### 2.1.3.2 Extração

A extração das características é uma das etapas que necessita de grande atenção, pois é responsável pela extração das características de cada objeto contido nas sentenças. Além disso, esta fase permite diferenciar sentimentos positivos e negativos em relação ao mesmo objeto, o que garante uma maior precisão no resultado final. (BERLIM, 2015).

A extração tem por objetivo identificar características do objeto sob análise que são alvo de comentários, como pares válidos a partir de um corpus representativo para o domínio escolhido. (SILVA, 2010).

Berlim (2015) salienta que:

É de extrema importância que sejam previamente selecionadas características que realmente sejam pertinentes para análise, reduzindo assim o número de características e extraído-se apenas as que serão interessantes e válidas. Em contrapartida é preciso ter o cuidado de não reduzir muito o número de características a serem analisadas para não diminuir o desempenho da extração. (BERLIM, 2015, p. 42).

### 2.1.3.3 Classificação

Na etapa de classificação, as sentenças subjetivas são classificadas de acordo com a sua polaridade, e as opiniões podem ser neutras, negativas ou positivas.

Para Berlin (2015, p. 43), a classificação é uma etapa fundamental no que diz respeito aos resultados, pois para se ter a resposta concreta em relação à entidade pesquisada, é preciso definir a polaridade da opinião, ou seja, se elas são positivas, negativas ou neutras.

#### 2.1.3.4 Sumarização

A sumarização é a última etapa do processo, e é responsável por agregar e representar os resultados da análise de sentimento. As informações obtidas a partir da classificação das sentenças são mostradas de diversas formas, desde textos descrevendo os resultados da análise a gráficos que fazem comparações entre dois ou mais produtos do mesmo domínio. (SILVA, 2010).

Liu (2012) alerta que, somente a opinião de uma pessoa não é suficiente para a realização da mineração de opinião. Na maioria das aplicações é necessário coletar um número expressivo de opiniões.

Liu (2012) ainda discorre que apesar da visualização das informações ser demonstrada de várias maneiras, a sumarização deve apresentar opiniões sobre diferentes entidades e seus aspectos, mostrando uma perspectiva quantitativa.

A fim de apresentar um possível campo para aplicação da análise de sentimento, a próxima sessão discorre sobre as redes sociais.

## 2.2 REDES SOCIAIS

As redes sociais referem-se a um conjunto de indivíduos e organizações, conectados, que vão construindo e reconstruindo a estrutura social. Essa conexão ocorre por meio das relações sociais que se manifestam de maneiras diversas e expressam a complexidade do mundo social. (PINTO; JUNQUEIRA, 2009).

Gabriel (2011) sobre as redes sociais discorre que:

Podem ser definidas como uma 'estrutura social formada por indivíduos (ou empresa), chamados de nós, que são ligados (conectados) por um ou mais tipos específicos de interdependência, como amizade, parentesco, proximidade/afinidade, trocas financeiras, ódios/antipatias, relações sexuais, relacionamento de crenças, relacionamento de conhecimento e de prestígio, etc.'. (GABRIEL, 2011, p.19).

Já para Recuero (2009, p. 24) uma rede social é definida como um conjunto de dois elementos: atores (pessoas, instituições ou grupos; os nós da rede) e suas conexões (interações ou laços sociais).

Este tipo de rede, passa a representar um conjunto de participantes autônomos, unindo ideias e recursos em torno de valores e interesses compartilhados. (MARTELETO, 2001).

Deve se levar em conta que redes sociais e mídias sociais são termos diferentes, em que as redes sociais apresentam-se como pessoas conectadas com interesses mútuos, as mídias sociais associam-se a conteúdos – textos, imagens – vídeos, etc. – gerados e compartilhados entre indivíduos participantes das redes sociais. Sendo assim, as redes e mídias sociais não são propriamente ditas uma tecnologia, mas sim pessoas e suas conexões. A tecnologia é somente um facilitador que favorece a interação entre os indivíduos e a geração ou compartilhamento de conteúdo. Dessa forma, sites de redes sociais, como é o caso do Facebook, Twitter, Youtube, entre outros, são as plataformas que tornam possível, facilitam e potencializam a conexão de pessoas com outras pessoas, ampliando em escala global o alcance das redes sociais. (GABRIEL, 2011).

A influência que as redes sociais possuem em relação a criação e compartilhamento de informações, justificam as análises que são realizadas nelas, principalmente a mineração de opinião, a qual é estudada neste trabalho. Marteleto (2001) discorre que:

Para estudar como os comportamentos ou as opiniões dos indivíduos dependem das estruturas nas quais eles se inserem, a unidade de análise não são os atributos individuais, mas o conjunto de relações que os indivíduos estabelecem através das suas interações uns com os outros. A estrutura é apreendida concretamente como uma rede de relações e de limitações que pesa sobre as escolhas, as orientações, os comportamentos e as opiniões dos indivíduos. (MARTELETO, 2001, p. 72).

Abaixo são abordadas duas redes sociais em que a mineração de opinião pode ser aplicada. Uma delas é o Twitter, a qual este trabalho utilizou como teste. A outra é o Facebook, também passível de aplicação da análise de sentimento.

### 2.2.1 Twitter

O Twitter foi lançado em março de 2006, nos Estados Unidos, após a parceria de Jack Dorsey com Biz Stone e Evan Williams. A ideia surgiu de Jack, durante uma sessão de *brainstorming* em que ele sugeriu um serviço de troca de status, o qual

também era o primeiro nome do serviço. O nome não representava o real significado da ferramenta e passaram a buscar em dicionários nomes parecidos com *twitch* – vibração do celular ao receber uma notificação, no caso da ferramenta uma mensagem curta a respeito de uma atualização – encontrando o termo Twitter que significa basicamente “uma pequena explosão de informações inconsequentes” e “pios de pássaros”. (TECMUNDO, 2010). Na FIGURA 5, é possível observar um perfil de um usuário da rede, no caso, a do próprio criador.

FIGURA 5 - PÁGINA DO PERFIL DE UM DOS CRIADORES DO TWITTER

The image shows the Twitter profile of Jack Dorsey (@jack). At the top, there are navigation icons for 'Início', 'Moments', 'Notificações', and 'Mensagens', along with a search bar and a 'Tweetar' button. The profile header includes a profile picture of Jack Dorsey, a banner image of a sunset, and statistics: TWEETS (22 mil), SEGUINDO (2.815), SEGUIDORES (4,06 mi), CURTIDAS (17,4 mil), and LISTAS (3). Below the header, there are tabs for 'Tweets', 'Tweets e respostas', and 'Mídia'. The main content area shows a tweet from Jack Dorsey (@jack) posted 16 hours ago, which reads: 'Thanks for the \$5-for-\$1 #EDFMatch @tfadell, @MegWhitman, and @nestmattl'. Below the tweet, there is a 'Traduzir do inglês' button and a quote: 'Jack, Thank you so much for your generous donation! Your gift to defend the climate will be matched \$5-for-\$1 by generous EDF donors like Meg Whitman, Tony Fadell, Matt Rogers, and @nestmattl'. To the right, there is a 'Quem seguir' section with a list of users to follow, including IndieGala, eremita, and livia. At the bottom right, there is a 'Assuntos do Momento' section with trending topics like #OneLoveManchester and #GoKeyd.

FONTE: Twitter (2017).

Esta ferramenta é uma rede social, porém com características únicas, como por exemplo a forma de publicação, que envolve o post de uma mensagem (Tweet) que deve conter no máximo 140 caracteres. Também possui uma funcionalidade chamada “*Trending Topics*”, que mostra os assuntos mais comentados pelas pessoas em cada região, além de trocas de mensagens privadas entre usuários e mensagens direcionadas a um usuário específico, através da citação no início da mensagem – técnica conhecida como endereçamento – utilizando o caractere “@”. Nos casos em que o tweets não são privados, eles são públicos e podem ser vistos por qualquer pessoa que realize uma busca na ferramenta. (NOGUEIRA, 2015). Permite ainda, que usuários sigam outros usuários, o que faz com que recebam todas as atualizações (tweets) deste usuário, sem a necessidade da pessoa seguida, também seguir seu

seguidor. Nesta ferramenta, os seguidores são conhecidos como “*followers*” e a ação de seguir um usuário chamada de “*following*”. (NOGUEIRA, 2015).

O Twitter disponibiliza dados para desenvolvedores através da sua API REST, os quais podem ser coletados através da utilização de metadados disponibilizados pelo Twitter. Metadados são descrições de dados armazenados em banco de dados, comumente definido como dados sobre dados a partir de um dicionário digital de dados. Tal dicionário normalmente é utilizado para organizar os metadados. A sua principal finalidade é documentar e organizar de maneira estruturada os dados das organizações com o objetivo de minimizar duplicação de esforços e facilitar a manutenção e recuperação dos mesmos. (SOUZA; CATARINO; SANTOS, 1997).

### 2.2.2 Facebook

O Facebook é um site e serviço de rede social que foi lançado em 4 de fevereiro de 2004, operado e de propriedade privada da empresa norte-americana Facebook Incorporation. O nome original dessa rede social foi Thefacebook. Foi fundado por Mark Zuckerberg e por seus colegas da Universidade de Harvard; Eduardo Saverin, Dustin Moskovitz e Chris Hughes. A estrutura e a utilização do site foi primeiramente limitada, pelos fundadores, aos estudantes da Universidade de Harvard, todavia foi expandida para outras faculdades da região onde se localizava a instituição de ensino. (VERAS, 2013). Em 27 de junho de 2017 o Facebook atingiu a marca de 2 bilhões de usuários em todo o mundo. (G1, 2017).

Os membros do Facebook, após realizarem um cadastro, podem criar um perfil, adicionar outros membros como amigos, seguir páginas, entrar em grupos, publicar postagens de outros sites, ou publicações com autoria própria, além de imagens, vídeos, notícias, etc. Cada postagem pode ser curtida, compartilhada ou comentada. (VERAS, 2013).

## 2.3 CRIPTOMOEDAS

As Criptomoedas são um tipo de dinheiro eletrônico que utilizam a criptografia em seus algoritmos para a geração de novas moedas ou para validar as transações. Para Ahamad, Nair e Varghese (2013) as Criptomoedas são arquivos físicos pré-

computados, que utilizam uma chave pública e uma chave privada geradas por um algoritmo de criptografia específico. A chave atribui a propriedade da moeda, a pessoa que está na posse da chave privada. Estas moedas utilizam a internet como meio de circulação.

O primeiro artigo que apresentou um possível sistema criptografado de pagamentos foi apresentado por David Chaum em 1982 e foi intitulado “*Blind Signatures for untraceable payments*”. Em 1990 Chaum, Fiat e Naor apresentam em seu artigo “*Untraceable Electronic Cash*” um protótipo de dinheiro eletrônico que buscava eliminar o problema de gasto duplo – falha de sistemas de criptomoedas que permite um usuário gastar a mesma moeda mais de uma vez – o qual chegou a ser implantado. (GARCIA, 2014).

## 2.4 BITCOIN

Nos próximos tópicos, é abordado em especial a criptomoeda Bitcoin, a qual, além de ser a mais utilizada entre as moedas virtuais, também é a mais conhecida.

### 2.4.1 O Que É Bitcoin

O Bitcoin é uma criptomoeda e um sistema de pagamento independente de qualquer autoridade central, em que seu valor ou oferta não é controlado por nenhum banco central, como ocorre com as demais moedas. (BRITO; CASTILLO, 2013).

Segundo Ulrich (2014, p. 15), “Bitcoin é uma forma de dinheiro, assim como o real, o dólar ou o euro, com a diferença de ser puramente digital e não ser emitido por nenhum governo. O seu valor é determinado livremente pelos indivíduos no mercado.” Outra definição trazida por Ulrich (2014) é que:

Bitcoin é uma moeda digital peer-to-peer (par a par ou, simplesmente, de ponto a ponto), de código aberto, que não depende de uma autoridade central. Entre muitas outras coisas, o que faz o Bitcoin ser único é o fato de ele ser o primeiro sistema de pagamentos global totalmente descentralizado. (ULRICH, 2014, p.17).

Esta moeda digital foi introduzida por Satoshi Nakamoto – pseudônimo de um usuário ou grupo que ainda é desconhecido – em seu artigo intitulado “Bitcoin: A Peer-

to-Peer Eletronic Cash System” publicado em outubro de 2008. Neste artigo Nakamoto (2008) mostra como deveria funcionar um sistema de pagamento totalmente eletrônico, baseado em um sistema criptografado que dispensa a participação de terceiros, um dos principais motivos para a criação do Bitcoin, pois utiliza a tecnologia par a par ou ponto a ponto. Neste artigo ainda é retratado o mecanismo geral de funcionamento, trazendo conceitos técnicos aplicados à infraestrutura de software e de rede.

Ulrich (2014) a respeito da divulgação do trabalho de Satoshi Nakamoto diz:

Satoshi tomou a iniciativa incrível de reinventar a moeda na forma de código de computador. O resultado foi o Bitcoin, introduzido ao mundo na forma menos promissora possível. Nakamoto lançou-o com um ‘white paper’ em um fórum aberto: aqui está uma nova moeda e um sistema de pagamento. Usem se quiserem. (ULRICH, 2014, p.12).

De acordo com o site Bitcoin.org (2017), não existe uma pessoa ou instituição que é dono da rede Bitcoin, em que a mesma é controlada por todos os usuários que fazem parte da rede em qualquer parte do mundo. Justamente por ser uma rede, o correto funcionamento do Bitcoin é dado por um consenso entre todos esses usuários.

No Bitcoin, transferências financeiras das moedas envolvem apenas os componentes tecnológicos de cada usuário, em que bits, contendo a codificação da moeda são transferidos de um cliente para outro cliente de modo irreversível. Todo esse processo é passível de verificação, pois ocorre publicamente e é armazenada no blockchain, conforme explica Ulrich (2014):

Todas as transações que ocorrem na economia Bitcoin são registradas em uma espécie de livro-razão público e distribuído chamado de blockchain (corrente de blocos, ou simplesmente um registro público de transações), o que nada mais é do que um grande banco de dados público, contendo o histórico de todas as transações realizadas. Novas transações são verificadas contra o blockchain de modo a assegurar que os mesmos Bitcoin não tenham sido previamente gastos, eliminando assim o problema do gasto duplo. A rede global peer-to-peer, composta de milhares de usuários, torna-se o próprio intermediário. (ULRICH, 2014, p.18).

O primeiro Bitcoin foi minerado no dia 3 de janeiro de 2009, por Satoshi Nakamoto, o qual resultou em uma quantidade de cinquenta Bitcoins, conforme pode ser verificado na FIGURA 6. O bloco zero foi chamado de “*The Genesis Block*”, por ser o bloco gerador da moeda.

FIGURA 6 - BLOCO 0 - THE GENESIS BLOCK

**Block #0**

BlockHash 00000000019d6689c085ae165831e934ff763ae46a2a6c172b3f1b60a8ce26f

### Summary

Number Of Transactions	1	Difficulty	1
Height	0 (Mainchain)	Bits	1d00ffff
Block Reward	50 BTC	Size (bytes)	285
Timestamp	Jan 3, 2009 4:15:05 PM	Version	1
Mined by		Nonce	2083236893
Merkle Root	4a5e1e4baab89f3a32518a88c31bc8...	Next Block	1

FONTE: Block Explorer (2017).

#### 2.4.2 Como surgem os Bitcoins

Conforme já mencionado anteriormente, o Bitcoin dispensa a participação da confiança em terceiros para que as transações sejam realizadas, pois é uma rede de peer-to-peer.

A produção de Bitcoins é determinada em seu código fonte, de modo que não é influenciado por mudanças externas e é produzido através da utilização de computadores programados para resolverem problemas matemáticos. (NAKAMOTO, 2008).

Silva (2014) a respeito do processamento matemático explica que:

Os mineradores, ao realizar o processamento do software Bitcoin, buscam encontrar uma sequência de dados (é a resposta do desafio, denominado 'nonce', na verdade, trata-se de um número que tem o seu valor influenciado pelo grau de dificuldade do bloco, definido pelo software básico do Bitcoin). Esta sequência de dados deve solucionar a seguinte situação: (nonce) + (hash do bloco anterior da cadeia do blockchain) + (hash do bloco a ser processado) = (hash final específico, que inicie com 12 zeros). (SILVA, 2014, p. 226).

Quando um minerador (que é um participante, ou nó da rede) encontra a sequência esperada, através de um processo denominado prova de trabalho, ele comunica com os demais nós, os quais validam a solução encontrada e adicionam este bloco ao blockchain (cadeia de blocos).

Como esses novos blocos descobertos sempre são adicionados a cadeia de blocos principal, ela contém todo os traços das transações, armazenados como um histórico. Dessa forma, o único jeito de alterar qualquer bloco é através de esforços

computacionais para solucionar todos os blocos que fazem parte da cadeia principal. Está é uma medida de segurança do Bitcoin, que torna impraticável a utilização desse processamento para refazer o processo de solução dos blocos, sendo mais vantajoso para os nós não honestos utilizar seus recursos para adquirir novas moedas, minerando de forma honesta. (NAKAMOTO, 2008).

O minerador (nó participante da rede) que encontrou a solução recebe uma quantidade de Bitcoins referente ao seu esforço – entende-se por esforço o gasto de energia e tempo de processamento para encontrar a solução. A quantidade de Bitcoins por bloco minerado varia com o decorrer do tempo, devido uma regra pré-estabelecida que a cada 210 mil blocos – cerca de quatro anos – esse número seja dividido pela metade (50%). A rede é programada para aumentar a oferta monetária como uma série geométrica até que o número total de Bitcoins chegue a 21 milhões. No momento em que este trabalho é escrito, a bonificação por bloco é de 12,5 Bitcoins.

De acordo com Ulrich (2014), estima-se que os mineradores colherão o último “satoshi”, ou 0,00000001 de um Bitcoin, no ano de 2140, e quando o fizerem, os mineradores direcionaram sua potência de processamento ao ato de verificar transações, e serão recompensados com taxas de serviço, em vez de novos Bitcoins minerados. Isso garante que os mineradores ainda tenham um incentivo de manter a rede operando após a extração do último Bitcoin.

#### 2.4.3 Transações De Bitcoins

A primeira transação de Bitcoins entre contas – de acordo com o site History of Bitcoin (2017) – ocorreu em janeiro de 2009, entre Satoshi e Hal Finney. A mesma está registrada no bloco 170 e foi realizada no valor de 10 BTC – BTC é a sigla mais utilizado para representar o Bitcoin.

No entanto a primeira transação do Bitcoin como meio de troca conhecida ocorreu em maio de 2010, quando “laszlo”, trocou uma pizza por 10 mil BTC. O fato é que, desde então, Bitcoins passaram a funcionar como meio de troca, de acordo com o seu objetivo fundamental. (ULRICH, 2014).

Ulrich (2014) explica de uma maneira resumida como as transações ocorrem:

As transações são verificadas, e o gasto duplo é prevenido, por meio de um uso inteligente da criptografia de chave pública. Tal mecanismo exige que a cada usuário sejam atribuídas duas 'chaves', uma privada, que é mantida em segredo, como uma senha, e outra pública, que pode ser compartilhada com todos [...]. A criptografia de chave pública garante que todos os computadores na rede tenham um registro constantemente atualizado e verificado de todas as transações dentro da rede Bitcoin, o que impede o gasto duplo e qualquer tipo de fraude. (ULRICH, 2014, p.18).

Para que seja possível a realização de transações, cada usuário da rede é identificado por uma chave pública, o que identifica ele na mesma, além de ser o endereço de sua carteira de Bitcoins. O endereço Bitcoin é um identificador único que possui letras e números, começando com o dígito 1 ou com o dígito 3, como por exemplo: 3FfmbHfpoiLjKFvyi1okTjJe2sN415paPI. Este tipo de endereço seria uma conta válida, capaz de receber e enviar moedas. Da mesma forma, este usuário possui uma chave privada, que representa a senha da carteira. Quando se possui estas duas chaves, torna-se possível a realização de transações. Uma transação na economia Bitcoin nada mais é do que um usuário escrever o endereço – a chave pública – do usuário que receberá os Bitcoins. A segurança está relacionada com a criptografia, em que a mensagem de envio é codificada utilizando a chave privada do usuário que possui as moedas para assinar com o endereço (chave pública) contemplado. Um ponto importante é que caso a chave privada associada a chave pública seja perdida, não será mais possível acessar aquele endereço, de modo que todas as moedas também sejam perdidas.

O próximo passo ocorre quando a transação é enviada para a rede, em que se espera a validação pelos nós, para que a mesma seja inserida em um bloco na cadeia de bloco (blockchain). No momento de validação é verificado o histórico do blockchain – através de um servidor de “carimbo de tempo” – para invalidar o gasto duplo. As transações são verificadas pelos próprios mineradores, que são recompensados por realizar essa validação.

Diferente de uma transação que passa por terceiros de confiança, como bancos e instituições financeiras, em que, se mantem a privacidade ao limitar o acesso das informações somente com as partes envolvidas na negociação, o Bitcoin registra todo o histórico de transações realizadas publicamente, de modo que é possível qualquer interessado verificar a quantidade enviada de uma pessoa a outra, sem informações que liguem essa transação a alguém. (NAKAMOTO, 2008).

#### 2.4.4 Características Do Bitcoin

De acordo com Nakamoto (2008) e Ulrich (2014) as principais características do Bitcoin quanto tecnologia são:

- a) Sistema distribuído: Não existe uma empresa ou dono que regule esta tecnologia, mas sim uma comunidade formada pelos usuários que fornece suporte e desenvolvimento, garantido o seu funcionamento. Quando o sistema é centralizado, caso ocorra falência ou alguma falha no sistema, tudo o que está envolvido, é afetado.
- b) Velocidade: Transações ocorrem de forma rápida e segura, além de funcionar em escala global de modo instantâneo, sendo influenciadas somente pela velocidade da internet das partes envolvidas.
- c) Custo de Transações: Um dos principais fatores que influenciam ou não na utilização de uma moeda, são os custos envolvidos em uma transação. No caso do Bitcoin as taxas são muito pequenas quando comparadas a outras formas de pagamentos e até o momento em que este trabalho é escrito, livres de taxas relacionadas a governos.
- d) Funcionalidade: A tecnologia Bitcoin funciona vinte e quatro horas por dia, todos os dias da semana e em qualquer lugar do mundo, necessitando unicamente de uma conexão com a internet. Essa característica está diretamente ligada com a forma descentralizada da tecnologia, em que, não é necessário a participação de um terceiro no processo, permitindo a sua utilização quando desejada.
- e) Transações são irreversíveis: Quando uma transação Bitcoin é efetuada, não é possível desfazer a mesma através de estorno ou alegando que a transação não foi efetuada por você mesmo.

Já o Bitcoin como moeda, ainda de acordo com Nakamoto (2008) e Ulrich (2014), possui as seguintes características:

- a) Caráter Pseudônimo ou anônimo: as transações de Bitcoin possuem registros na blockchain, onde estão armazenadas todas as transações realizadas até hoje. No entanto, não existe uma identificação dos usuários

que realizaram essas transações, somente relações de seu endereço (chave pública).

- b) Escassez: Existe um número máximo de Bitcoins para serem utilizados como unidade da moeda, limitado matematicamente a 21 milhões de BTC.
- c) Fungibilidade e divisibilidade: Um único Bitcoin pode ser dividido em frações menores, de até 8 casas decimais. Todas essas frações continuam a ser equivalentes e podem ser transmitidas sem distinções.
- d) Mercado Livre: Taxa de cambio e poder de compra são determinados pelo mercado de acordo com a lei de oferta e procura – para o Bitcoin, a oferta é pré-determinada e previsível, pois está definida no protocolo, porém a demanda varia de acordo com a situação da economia global, velocidade de adoção da tecnologia, legislação e moedas concorrentes - não existindo uma instituição que regulamente o valor que deve ser atribuído a um Bitcoin quando o mesmo for trocado por outra moeda, produto ou serviço.

#### 2.4.5 Bitcoin no Brasil Hoje

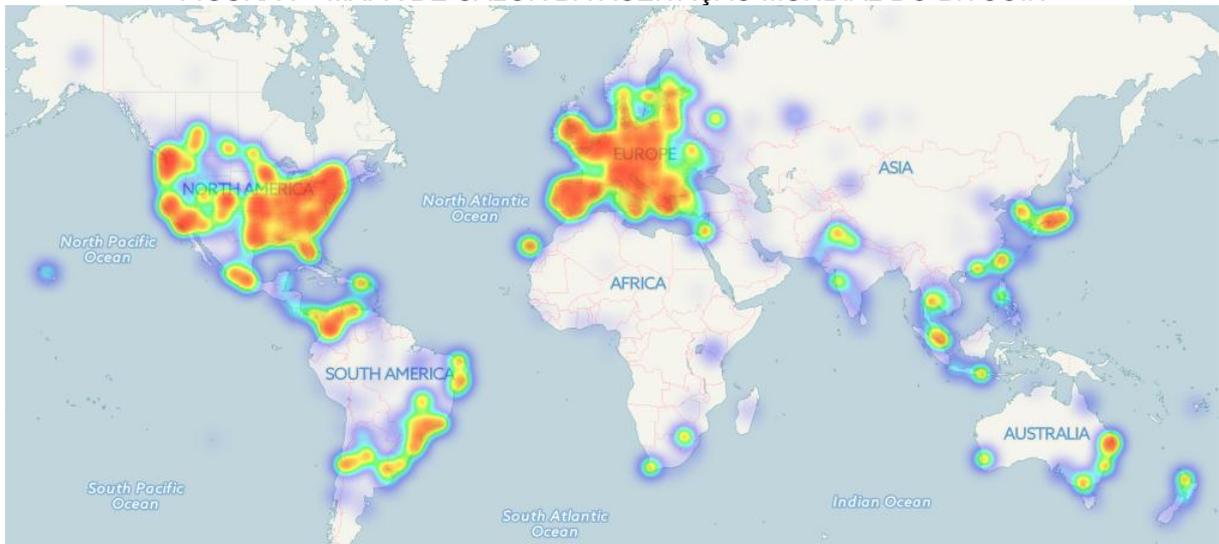
Muito vem sendo discutido a respeito do Bitcoin na sociedade brasileira atualmente, principalmente pela sua forma de investimento de risco. De acordo com o site BitcoinBrasil (2017), no ano de 2015, as incertezas na economia e também no cenário político brasileiro, fizeram com que investidores avaliassem outras maneiras para aplicar seu dinheiro. E quem escolheu pelo uso da moeda digital, teve uma valorização de cerca de 4809% caso comparado com a cotação atual do Bitcoin no momento em que este trabalho é escrito – R\$ 42899,00 – e o início do ano de 2015 – R\$892,00.

Outro fato que está fazendo com que a moeda se torne realidade no país é o das empresas passarem a aceitar Bitcoin como moeda de compra por produtos ou serviços, em que bares, lojas virtuais, pousadas, consultórios médicos e muitos outros ramos, estão proporcionando uma nova maneira de pagamento para seus clientes, além de se beneficiarem com as taxas baixíssimas de transação. Outro exemplo de empresa que passou a utilizar o Bitcoin é a Tecnisa – empresa do mercado imobiliário brasileiro – que possibilita o uso de Bitcoins como parte do pagamento de um imóvel.

Além de aceitar, a empresa garante mais um bônus de 5% na primeira parcela. (TECNISA, 2017).

É possível verificar os estabelecimentos que aceitam Bitcoin no mundo através de um mapa colaborativo chamado Coinmap (2017). Este mapa mostra estabelecimentos físicos e comerciais que aceitam a moeda digital e onde estão localizados. A FIGURA 7 apresenta um mapa de calor que representa a aceitação mundial do Bitcoin.

FIGURA 7 - MAPA DE CALOR DA ACEITAÇÃO MUNDIAL DO BITCOIN



FONTE: Coinmap (2017).

No Brasil existem também alguns caixas eletrônicos de Bitcoin, um deles, instalado no Centro Empresarial Nações Unidas (Cenu), localizado na Zona Sul de São Paulo, já chegou a movimentar R\$180.000 em transações nos dois primeiros meses de uso. (VEJA, 2017). O caixa eletrônico mencionado acima, permite compra e venda de Bitcoin além do saque em reais, diferenciando-se de outros caixas que só permitem a compra do Bitcoin através da inserção de cédulas. Com a mesma opção de troca de reais para Bitcoin ou vice-versa, a empresa Bitcoin to you, possui algumas lojas físicas que funcionam como casas de câmbio, o que faz com que a moeda digital não fique restrita a internet. (G1 2014).

No contexto governamental, a receita federal, desde 2014, diz que Bitcoins se equiparam a ativos financeiros para fins tributários, dessa forma devem ser declarados no imposto de renda. Para pessoas que possuem valores superiores a R\$1.000, devem declarar a moeda como outros bens e impostos de 15% devem ser pagos nas

vendas de valores superiores a R\$35 mil (FOLHA DE SÃO PAULO, 2014). O Banco Central do Brasil (2014), em comunicado do dia 19 de fevereiro de 2014, analisou os riscos relacionados com a aquisição de moedas virtuais e de suas transações, e classificou-se como acompanhante dos instrumentos, a fim de poder adotar medidas que pudessem ser eventualmente necessárias.

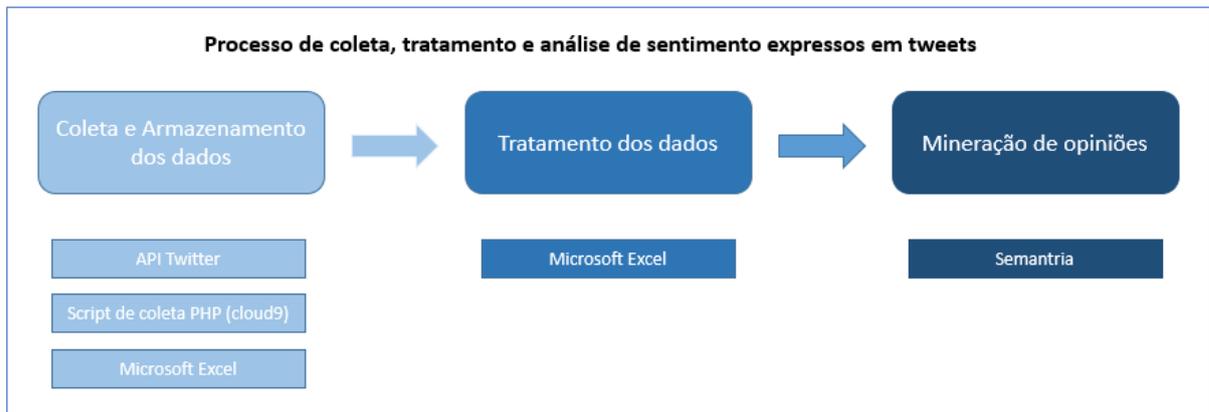
No quesito legislativo, o Deputado Federal Aureo Ribeiro, propôs o Projeto de Lei nº 2303/2015, o qual possui a seguinte ementa “Dispõe sobre a inclusão das moedas virtuais e programas de milhagem aéreas na definição de "arranjos de pagamento" sob a supervisão do Banco Central”. O projeto de lei não fala a respeito do Bitcoin em si, porém aborda o tema moedas virtuais, de modo que o mesmo estaria envolvido com a legislação. Até o momento em que este trabalho está sendo escrito, o projeto de lei está com a situação de Aguardando Constituição de Comissão Temporária pela Mesa.

Visto esta breve descrição do Bitcoin, espera-se realizar uma análise de opiniões expressas na rede social Twitter sobre esta moeda.

### 3 PROCEDIMENTOS METODOLÓGICOS

Neste capítulo estão descritos os procedimentos metodológicos tomados na execução deste trabalho. Os procedimentos estão divididos em etapas, mostradas na FIGURA 8. Compreendem a coleta e armazenamento dos dados, tratamento dos dados, mineração de opiniões e ferramentas.

FIGURA 8 - PROCESSO DE COLETA, TRATAMENTO E ANÁLISE DE SENTIMENTO EXPRESSOS EM TWEETS.



FONTE: O Autor (2017).

#### 3.1 DEFINIÇÃO DA POPULAÇÃO

O Twitter possui uma quantidade imensurável de informações – com uma média de 58 milhões de tweets por dia – e um potencial enorme como objeto de estudo para a mineração de opiniões, visto que possui cerca de 695 milhões de usuários registrados, dos quais 342 milhões são ativos (STATISTICS BRAIN, 2017). Berlim (2015) afirma que usuários do Twitter frequentemente utilizam o mesmo para expressar suas opiniões, a qual é uma ferramenta de interessante interface, além de ser uma rede de relacionamentos e interatividade que permite a geração de estratégias para a tomada de decisão. No entanto, a análise manual destas informações, mostra-se inviável, devido a imensa quantidade existente, o que se faz necessário a utilização de ferramentas que verificam se um tweet expressa positivamente ou negativamente uma opinião ou sentimento sobre determinado assunto ou atitude do usuário.

Desta maneira, o presente trabalho teve como população a rede social Twitter, a qual foram coletados os dados para a mineração de opinião. A Coleta dos dados

(seção 3.2) foi realizada através de um script criado pelo autor deste trabalho na linguagem PHP, que possui os parâmetros de busca da API REST do Twitter.

## 3.2 COLETA E ARMAZENAMENTO DOS DADOS

A etapa de coleta de dados compreende o recolhimento de registros para análise. No caso deste trabalho, a coleta foi feita exclusivamente na rede Twitter através de sua API denominada REST (seção 3.2.1).

### 3.2.1 API REST do Twitter

A sigla API corresponde ao termo “Application Programming Interface” que traduzido do inglês significa Interface de Programação de Aplicativos. Esta API nada mais é do que parâmetros e comandos de programação que garantem construção ou acesso a um determinado aplicativo (software) ou uma plataforma web. As empresas disponibilizam APIs, quando tem a intenção de que outros criadores de softwares desenvolvam produtos associados ao seu serviço. Em diversos casos, estas APIs são disponibilizadas junto de uma documentação, que facilita a integração das aplicações. De modo geral, a API é composta de uma série de funções acessíveis somente por meio de programação. (CANALTECH, 2017).

O Twitter possui um site destinado a explicação das ferramentas que se integram a rede social. Para este trabalho foi utilizado a API REST (Representantion State Transfer), a qual permite através da programação ler e escrever dados no Twitter. Além disso, é possível criar um novo tweet, ler perfis de usuários, como também alguns de seus dados. O padrão REST se baseia nas seguintes requisições HTTP: POST, PUT, GET e DELETE. (TWITTER DEVELOPERS, 2017).

A API REST possibilita chamadas em intervalos de 15 minutos, em que, após decorrer este tempo, as taxas limites de cada método são zeradas. Foi utilizado o método “GET search/tweets” neste trabalho, o qual, pode ser executado 180 vezes no intervalo de 15 minutos.

Para utilizar a API REST do Twitter como usuário, é necessário utilizar o protocolo OAuth, que é um protocolo de autorização para APIs. Através desta autorização, as aplicações conseguem acessar os recursos protegidos em nome de

um usuário. No caso do Twitter são geradas chaves e Tokens que devem ser informados na aplicação para permitir o acesso a conta. Existem outros métodos para garantir o acesso, porém não serão abordados neste trabalho.

### 3.2.2 Criando a aplicação

A primeira etapa para coletar os dados no Twitter é a criação de uma aplicação. Para isso, é necessária uma conta ativada na própria rede social, a qual possibilitará a execução da aplicação. Para tal, foi acessado o site <https://apps.twitter.com>, e criado um aplicativo preenchendo os dados conforme mostrado na FIGURA 9.

Para este trabalho, somente nome, descrição e website foram preenchidos, além de ser obrigatório concordar com os termos de desenvolvimento. O campo website foi preenchido com a própria ferramenta utilizada para o desenvolvimento do script, no caso a Cloud 9 – a mesma é descrita na seção 3.5 – visto que não existe a necessidade de disponibilização da aplicação, em que a mesma somente foi utilizada para fins de pesquisa.

Após ser criada a aplicação, foi gerado as chaves e tokens utilizados na autenticação da aplicação. Para isso foi acessado a aba “Keys and Access Tokens”. As chaves já estarão disponíveis no campo “Application Settings”, no entanto, os Tokens devem ser gerados no campo “Your Access Tokens”, clicando no botão “create my access token”. A FIGURA 10 apresenta a maneira como as chaves e tokens são mostrados para o usuário. Na imagem mencionada acima, tais informações foram mescladas por questões de privacidade, visto que referenciam a conta do Twitter do autor deste trabalho.

FIGURA 9 - INFORMAÇÕES NECESSÁRIAS PARA CRIAR UMA APLICAÇÃO

## Create an application

### Application Details

**Name \***

*Your application name. This is used to attribute the source of a tweet and in user-facing authorization screens. 32 characters max.*

**Description \***

*Your application description, which will be shown in user-facing authorization screens. Between 10 and 200 characters max.*

**Website \***

*Your application's publicly accessible home page, where users can go to download, make use of, or find out more information about your application. This fully-qualified URL will be shown in user-facing authorization screens. (If you don't have a URL yet, just put a placeholder here but remember to change it later.)*

**Callback URL**

*Where should we return after successfully authenticating? OAuth 1.0a applications should explicitly specify their oauth\_callback URL on the request token step, regardless of whether your application uses OAuth 1.0a. If you don't use OAuth 1.0a, leave this field blank.*

### Developer Agreement

Yes, I have read and agree to the [Twitter Developer Agreement](#).

Create your Twitter application

FONTE: Twitter (2017).

FIGURA 10 - CHAVES E TOKENS DA APLICAÇÃO

## Application Settings

*Keep the "Consumer Secret" a secret. This key should never be human-readable in your application.*

Consumer Key (API Key)	[REDACTED]
Consumer Secret (API Secret)	[REDACTED]
Access Level	Read and write (modify app permissions)
Owner	[REDACTED]
Owner ID	[REDACTED]

## Application Actions

Regenerate Consumer Key and Secret    Change App Permissions

## Your Access Token

*This access token can be used to make API requests on your own account's behalf. Do not share your access token secret with anyone.*

Access Token	[REDACTED]
Access Token Secret	[REDACTED]
Access Level	Read and write
Owner	[REDACTED]
Owner ID	[REDACTED]

FONTE: Twitter (2017).

### 3.2.3 Executando a aplicação

O próximo passo é a aplicação do método de busca da API. Para este trabalho como já mencionado acima, foi utilizado o método GET search/tweets. Este método retorna uma coleção de tweets relevantes de acordo com parâmetros definidos pela aplicação. Para executar tal método é necessário a autenticação OAuth e o resultado da consulta vem no formato JSON (JavaScript Object Notation) – formato leve de troca de dados entre sistemas. Um exemplo de retorno de consulta é mostrado no apêndice A.

No entanto, este método não retorna todos os tweets sobre o assunto pesquisado, limitando-se a 100 registros por consulta. É possível realizar 180 consultas dentro do intervalo de 15 minutos, porém ao realizar mais de uma consulta

em um intervalo pequeno de tempo, a quantidade de tweets repetidos retornados é muito elevada, o que prejudica a questão de univocidade esperada nesta pesquisa. Desta maneira, foi realizada uma consulta diária somente (100 retornos/dia), durante 31 dias, o qual resultou nos 3063 tweets esperados para tratamento e seleção de 350 tweets únicos para análise.

O método GET search/tweets, possui 10 parâmetros de consulta, dos quais, somente 5 foram relevantes para este trabalho. Os parâmetros estão descritos no QUADRO 1.

QUADRO 1 - PARÂMETROS UTILIZADOS DO MÉTODO GET SEARCH/TWEETS

continua

<i>Nome parâmetro</i>	<b>Descrição</b>	<b>Valor utilizado</b>
<i>Q</i>	Este parâmetro é o valor que será pesquisado e é o único obrigatório. Pode ter até 500 caracteres junto com operadores.	'Bitcoin' -filter:retweets
<i>Lang</i>	Restringe a consulta a um idioma.	Pt
<i>result_type</i>	Especifica o tipo de resultado da pesquisa. Valores aceitos: "popular", "recent" e "mixed"	Recente
<i>Count</i>	O número de Tweets retornados por consulta. O valor máximo para este parâmetro é 100	100
<i>Until</i>	Retorna os tweets criados antes de uma data determinada. O formato da data deve estar no padrão YYYY-MM-DD. Este campo deve possuir datas dentro em um intervalo máximo de 7 dias em relação a data atual, pois os índices de busca do Twitter, retornam dados de no máximo 1 semana.	Este campo varia de acordo com a data a ser pesquisada. Caso a pesquisa fosse realizada por exemplo na data 04/06/2017, o parâmetro receberia como valor: "2017/06/04". Dessa forma iria retornar tweets do dia 03/06/2017.

FONTE: O Autor (2017).

No parâmetro "q" foi utilizado o termo "'Bitcoin' -filter:retweets", de modo que, buscou-se coletar somente tweets que possuíam o termo Bitcoin, mas que não fossem um retweet – pelo motivo de que o mesmo já pode ter sido coletado anteriormente. O campo "lang" está restrito ao idioma PT (português), pois a mineração de opiniões foi realizada de acordo com o cenário brasileiro, restringindo a coleta no idioma oficial do Brasil. Para o campo "result\_type" foi definido o valor "recent", de modo a buscar os últimos tweets postados, objetivando novamente eliminar a repetição. O parâmetro

“count” foi definido com o valor “100”, pois como default, o método retorna somente 15 registros. O parâmetro “until”, foi utilizado em alguns casos em que ocorreram problemas na coleta no dia (esquecimento, falta de conexão, etc.), de modo que o mesmo permite estipular a coleta de registros antes de uma data definida.

O script capaz de realizar a consulta e a coleta de tweets foi desenvolvido na linguagem PHP (acrônimo recursivo para PHP: Hypertext Preprocessor) – linguagem de script open source de uso geral, muito utilizada, especialmente adequada para o desenvolvimento web e que pode ser embutida dentro do HTML – pelo prévio conhecimento básico do autor deste trabalho da linguagem. O Script criado pode ser visualizado no apêndice B.

Para evitar que o script retorne todos os campos indexados de um Tweet – conforme mostrados no retorno padrão no apêndice A – no próprio script foi criada uma função que retorna somente os metadados utilizados na mineração de opinião ou na análise. Esta função determina que um registro que equivale a um tweet seja separado pelo marcador “;” no final de cada metadado e por fim quebrado pela tag html “<br>”. Foi realizado desta forma a fim de facilitar a importação dos dados para a ferramenta Excel. Os metadados referentes a publicação estão descritos no QUADRO 2.

QUADRO 2 - METADADOS COLETADOS QUE REFERENCIAM A PUBLICAÇÃO

<b>Metadado</b>	<b>Descrição</b>	<b>Exemplo</b>
<i>created_at</i>	Horário em que a publicação foi criada. (Formato UTC)	Sat Jun 03 23:59:44 +0000 2017
<i>id_str</i>	Identificador único da publicação	871154467101696001
<i>Text</i>	O texto contido dentro da publicação. Este campo é o texto do tweet. (Formato UTF-8)	Quando eu crescer quero ser farmador de Bitcoin
<i>Lang</i>	Corresponde a linguagem detectada	Pt
<i>favorite_count</i>	Informa o número de curtidas que uma publicação recebeu	3
<i>retweet_count</i>	Informa o número de vezes que um tweet foi retweetado.	2
<i>quoted_status_id_str</i>	Informa o ID do Twitter que está sendo citado	871105228837330944

FONTE: O Autor (2017).

O metadado “created\_at” foi obtido para controle da coleta diária dos tweets. O metadado “id\_str” foi coletado pois representa o identificador único daquele tweet e

utilizado no tratamento dos dados em relação a tweets repetidos, os quais foram excluídos. O metadado “text” é o mais importante, pois contém o tweet escrito pelo usuário, o qual foi utilizado na análise de opiniões. O campo “lang” foi extraído para validar se a API está retornando ou não os registros somente no idioma português. Os metadados “favorite\_count” e “retweet\_count” foram coletados para verificar a aceitação daquela tweet, pois o primeiro representa o número de curtidas que o mesmo recebeu de outros usuários da rede e o segundo a quantidade de vezes que aquele tweet foi retweetado. O atributo “quoted\_status\_id\_str” foi obtido e validado com o campo “id\_str” em busca de tweets repetidos. Quando uma citação é realizada, o tweet é copiado, perdendo a univocidade esperada para análise.

Para realizar o armazenamento dos registros coletados, foi utilizado o editor de planilhas Microsoft Excel, o qual é descrito na seção 3.5. Primeiramente, após ser executado a extração dos dados via API, o resultado é mostrado em uma página HTML criada junto com o script. Os dados desta página foram então copiados e colados em um bloco de notas e salvos no formato CSV – formato de arquivo que armazena dados tabelados – neste caso separados pelo delimitador ponto e vírgula (;) conforme mencionado anteriormente. O resultado de uma consulta pode ser visualizado na FIGURA 11.

FIGURA 11 - RETORNO DE UMA CONSULTA REALIZADA COM O SCRIPT CRIADO

```
Sat Jun 03 21:01:24 +0000 2017;871109587025264640;Câmara dos Deputados quer tributar e regulamentar transações com Bitcoin. https://t.co/pYqfB4xqu;pt;0;0;;;Anderson Lima;AndersonLimaHW;sao paulo ;20458;2022;635;<br>
```

FONTE: O Autor (2017).

Em uma planilha do Excel, foi reservado a primeira linha para preencher com os metadados coletados, cada um separado por colunas. Ao todo foram utilizadas 7 colunas. Para um melhor entendimento, os metadados serão traduzidos para o português, sendo dispostos na seguinte sequência: Data, ID\_Tweet, Texto\_Tweet, Idioma\_Tweet, Numero\_likes, Vezes\_retweetado, ID\_tweet\_citado,

Os dados precisavam ser exportados diariamente após a realização da coleta. Para isso, foi utilizado a opção “Obter dados Externos de Textos” disponível na aba “Dados do Excel”. Para o tipo de campo foi utilizado “delimitado”, com o delimitador ponto e vírgula (;).

Foi identificado um problema após finalizado o processo de importação. A existência do caractere ponto e vírgula (;) no atributo “Text”, o que causou a quebra

dos dados de forma errônea. Ao todo, trinta e dois registros foram afetados, iniciando sempre na coluna Texto\_Tweet, e sobrepondo a ordem das demais colunas. Os registros em que ocorreram o caso mencionado, foram excluídos. A base ficou com 3031 registros.

### 3.3 TRATAMENTO DOS DADOS

A fim de evitar problemas relacionados a codificação dos resultados, foi acrescentado no próprio script, a tag meta – instruções em linhas de código HTML ou "etiquetas" que descrevem o conteúdo do site para a renderização de um navegador – para que as consultas e as respostas fossem realizadas na codificação UTF-8 – codificação binária (Unicode) de comprimento variável – utilizada pelo Twitter para armazenar os tweets em seus bancos de dados. A FIGURA 12 apresenta um exemplo da tag meta definindo a codificação para UTF-8.

FIGURA 12 - EXEMPLO DE DEFINIÇÃO DE CODIFICAÇÃO UTF-8 NO HTML

```
1 <!DOCTYPE html>
2 <html >
3 <header>
4   <meta charset="utf-8" />
5 </header>
6 <body>
```

FONTE: O Autor (2017).

Após ser realizado a última importação dos dados obtidos dentro dos trinta dias de coleta, certas modificações foram realizadas na base final.

A primeira delas foi converter o formato dos campos que armazenam os IDs das publicações (ID\_Tweet, ID\_tweet\_citado) de geral para número, removendo ainda as duas casas decimais. Este procedimento foi realizado para facilitar que o localizador do Excel encontre os IDs facilmente.

O campo ID\_tweet\_citado foi utilizado para verificar a presença de tweets repetidos. Nos casos em que o valor presente no atributo ID\_tweet\_citado era o mesmo do valor presente no atributo ID\_Tweet, foi removido todo o registro do ID\_tweet\_citado. Após ser realizado a verificação dos tweets\_citados a base ficou com 2914 registros.

O campo Idioma\_Tweet foi utilizado para validar se o script só retornou tweets do idioma português. Neste trabalho, todos os resultados foram retornados com o valor “pt”.

A etapa final e mais demorada do processo de tratamento dos dados, foi a de validação da coluna Texto\_Tweet, a qual deveria conter somente tweets únicos e os mesmos não deveriam ser spams. Para tal foi utilizado a opção filtro do Excel e o processo realizado manualmente. Primeiramente foi removido todos os textos que possuem atualizações do valor do Bitcoin, ou de outras moedas, mas que utilizam a palavra Bitcoin no texto do tweet por exemplo. Também foram removidos os tweets que possuíam URLs sem informações relevantes nos textos. Tweets com mensagens de como ganhar Bitcoins, minerar gratuitamente, ou possíveis fraudes, também foram removidos

Após a remoção dos possíveis SPAMs, foi realizado novo filtro em busca de tweets repetidos. Nos casos encontrados, os mesmos foram removidos.

Ao ser finalizado a etapa de seleção dos textos, 350 registros mostraram-se aptos para serem analisados. As devidas explicações do processo de análise de opinião são apresentadas na próxima seção.

### 3.4 MINERAÇÃO DE OPINIÃO

Para realizar a análise de opinião em um texto, é necessário realizar as etapas de subjetividade, extração, classificação e sumarização. Para este trabalho, foi utilizado o Software Semantria, pelo motivo da mesma rodar dentro da ferramenta Microsoft Excel e pela capacidade de analisar sentenças no idioma português. Além disso, este software proporciona a análise de opinião em todas as etapas, detalhando todos os níveis – documento, sentença e aspecto – e as respectivas polaridades. (BERLIM, 2015). Existem diversas outras ferramentas e técnicas, no entanto, não foram atendidas neste trabalho. A explicação da ferramenta é apresentada na seção 3.5.

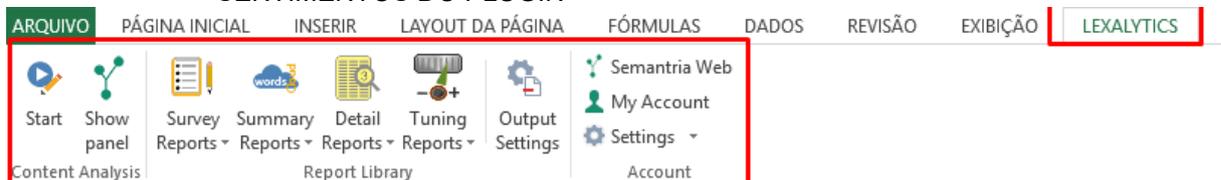
Conforme já mencionado, o software Semantria possui uma aplicação baseada em um plugin do Microsoft Excel, o qual foi utilizado neste estudo. Os procedimentos para instalação deste plugin estão descritos na seção 3.4.1 e a utilização do mesmo para a mineração de sentimentos na seção 3.4.2.

### 3.4.1 Instalação do Plugin Semantria para Excel

Para criar um cadastro que permita a utilização da ferramenta em sua versão trial, foi necessário utilizar um e-mail corporativo ou educacional, não sendo permitido a utilização de domínios como Gmail, Yahoo, Hotmail, entre outros. Ao finalizar o cadastro, é enviado um e-mail de confirmação que redireciona para a página de download do plugin. Concluindo o download, basta executar o pacote de instalação e aceitar os termos de uso, seguindo as etapas da instalação. Alguns componentes serão instalados, basta prosseguir e concluir a instalação.

Na sequência, ao executar o Excel, uma janela solicitando as credencias de utilização do plugin é mostrada. Estas credencias são enviadas junto ao e-mail de confirmação de cadastro. Ao inserir as informações solicitadas, o login é autenticado. Depois da autenticação, é possível notar a presença de uma nova aba a direita dos menus padrões do Excel. A FIGURA 13 mostra a aba “lexalytics” e suas opções, a qual é responsável por toda interação na realização da mineração de sentimentos.

FIGURA 13 - ABA LEXALYTICS RESPONSÁVEL PELAS OPÇÕES DE MINERAÇÃO DE SENTIMENTOS DO PLUGIN



FONTE: O Autor (2017).

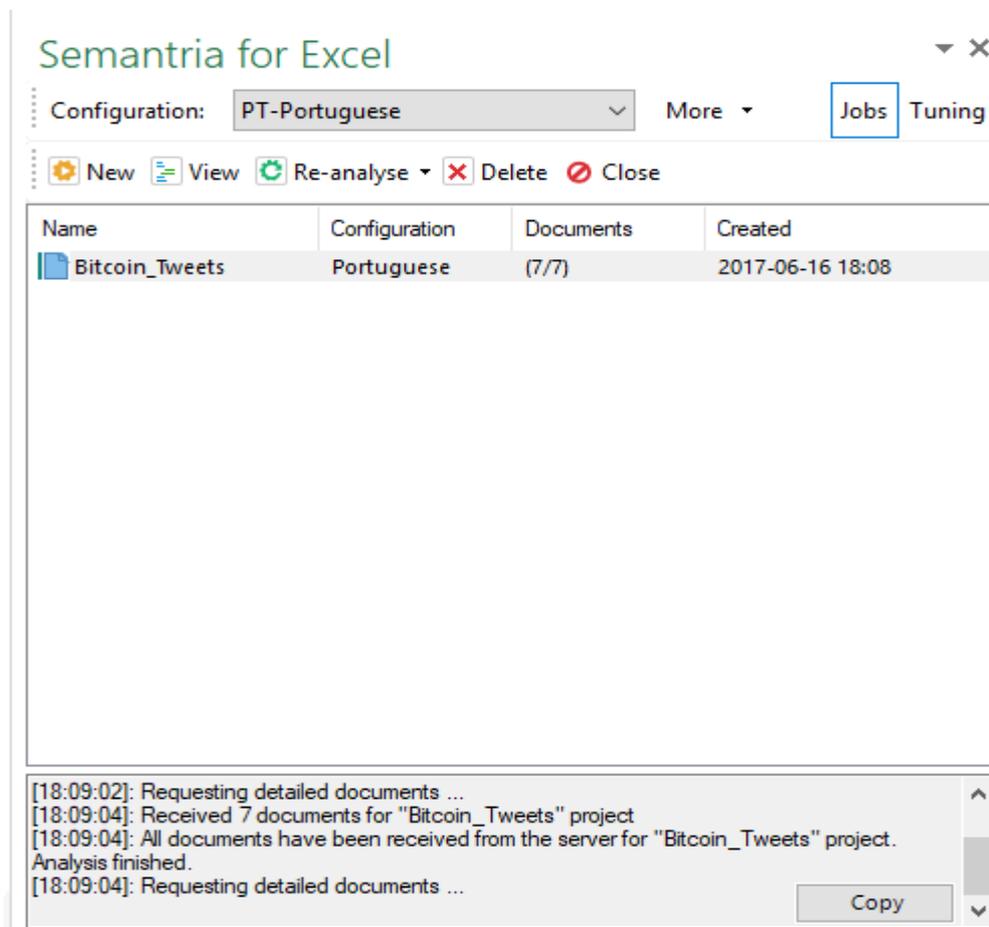
### 3.4.2 Etapas da mineração de opiniões com o Plugin Semantria para Excel

O plugin Semantria possui diversas utilidades, no entanto, para este trabalho somente foi utilizado a análise de sentimentos. Para iniciar a mineração de opiniões é necessário carregar a base de dados. Neste caso, foi utilizada a planilha com os registros tratados na etapa anterior.

Ao ser carregado a base, foi acessado a aba “lexalytics” e selecionado a opção “Start”, dentro do campo “Content Analysis”. Uma janela é aberta, a qual nada mais é do que um assistente para a realização da análise. A primeira etapa é definir a fonte de dados, no caso desta pesquisa, foi escolhido a opção “current worksheet” (planilha atual). Na sequência foi habilitado a opção “First row has column heading” (primeira linha possui cabeçalho), e definido as colunas dentro de três categorias –

Identifier Column, Text Columns e Metadata Columns. Na categoria “Identifier Column” foi definido o atributo ID e para a categoria “Text columns” o atributo Texto\_Tweet. As demais colunas foram definidas na categoria “Metadata columns”. Na próxima etapa foram definidas quais as linhas que iriam ser verificadas e o idioma do texto a ser analisado. Deve ser informado um nome para a análise executada. No último passo, foi selecionado a opção “Analyze” e o processo envolvendo todas as etapas da mineração de opiniões foi realizado. Ao ser finalizado o processamento, uma janela contendo os registros processados aparece, conforme FIGURA 14, indicando que o processo foi finalizado.

FIGURA 14 – PAINEL DE OPÇÕES DO PLUGIN SEMANTRIA



FONTE: O Autor (2017).

A próxima etapa consiste em selecionar o tipo de relatório para visualizar os registros processados. Para esta pesquisa, somente foi utilizado o relatório detalhado, denominado “Document Overview” (Visão geral do documento), que apresenta a polaridade dos registros conforme a FIGURA 15.

FIGURA 15 - ANÁLISE REALIZADA COM O PLUGIN SEMANTRIA

Docum ent ID	Sta tus	Source Text	Summary	Detected Language	Detected Language Score	Document Sentiment	Docume Sentime +/-
1	PR	eu fico desgraçado da cabeça como q	eu fico desgraçado da	Portuguese	0,5410824	0,13333334	neutral
2	PR	. @izzynobre vc nao acha melhor tirar	@izzynobre vc nao acha	Portuguese	0,7636435	0,69999999	positive
3	PR	@Bl3nd3r_TTL vendeu meio bitcoin ,	@Bl3nd3r_TTL vendeu meio	Portuguese	0,99999386	0	neutral
4	PR	@zuinix Após comprar o bitcoin qual	@zuinix Após comprar o	Portuguese	0,718532	0,40000001	positive
5	PR	@lucasduarty Você compra bitcoin	@lucasduarty Você compra	Portuguese	0,2798659	0	neutral
6	PR	@lucasduarty Eu n sei de corretora	@lucasduarty Eu n sei de	Spanish	0,24595462	0,69999999	positive
7	PR	Estou comprando BitCoin.	Estou comprando BitCoin...	Portuguese	0,9679125	0,60000002	positive

FONTE: O Autor (2017).

Este relatório é considerado a última etapa do processo de mineração de opinião, chamado de sumarização, em que as opiniões são apresentadas em formas de textos, gráficos ou comparações.

### 3.5 FERRAMENTAS NECESSÁRIAS

Este tópico traz uma visão geral sobre as principais ferramentas utilizadas na realização deste estudo.

A primeira delas é a Cloud9, um ambiente de desenvolvimento integrado totalmente online, que suporta diversas linguagens de programação, incluindo C, C++, PHP, Ruby, Perl, Python, JavaScript, entre outras. Esta ferramenta permite que desenvolvedores codifiquem imediatamente, com espaços de trabalhos configurados para a linguagem escolhida, além de permitir a execução de testes e execução na própria ferramenta. A ferramenta é gratuita, porém possui planos com certos diferenciais. No entanto, para a realização desta pesquisa, a versão gratuita atendeu perfeitamente as necessidades, visto que foi executado somente o script de a coleta de tweets.

Outra ferramenta utilizada, é o editor de planilhas Microsoft Excel, que possui inúmeras aplicações, além de uma interface intuitiva, com ferramentas de construções de gráficos e realizações de cálculos matemáticos. Neste trabalho, este software foi utilizado para armazenamento dos dados coletados pelo script de busca e para o tratamento dos dados.

Para analisar as opiniões coletas no twitter a respeito do Bitcoin, foi utilizado a ferramenta Semantria. Esta ferramenta é uma solução desenvolvida pela empresa Lexalytics, companhia conceituada no desenvolvimento de softwares de análise de

texto, é uma ferramenta de análise de sentimentos que pode ser aplicada a qualquer tipo de documentos, sentenças ou frases. A ferramenta possui um plugin, que é instalado no Microsoft Excel, a qual possibilita a análise e classificação dos sentimentos em positivos, negativos ou neutros. (BERLIM, 2015).

## 4 APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS

Esta seção apresenta a análise dos resultados obtidos durante a realização da pesquisa teste na rede social Twitter. Apresenta também um quadro com as facilidades e dificuldades encontradas pelo autor deste trabalho na execução dos procedimentos relatados no tópico 3.

### 4.1 ANÁLISE DA ETAPA DE SUMARIZAÇÃO

A última etapa do processo de análise de sentimentos é a sumarização dos resultados. Neste trabalho, tal etapa será discutida e apresentada através de textos e gráficos.

Vale salientar, que o objetivo deste trabalho não é apresentar e discorrer sobre os resultados propriamente ditos, de modo que, somente serão mostrados com o intuito de proporcionar uma base de entendimento para as dificuldades e facilidades encontradas no processo (seção 4.2).

Tais resultados, devem ser encarados apenas como um indicador e não como verdade absoluta, de modo que existem palavras e expressões nos tweets que a ferramenta não é capaz identificar, tornando assim errônea a classificação. Sem contar que gírias e ironias são frequentes na linguagem escrita da internet, atreladas com diversos erros de digitação ou abreviação de palavras, que acabam prejudicando a execução do algoritmo da ferramenta.

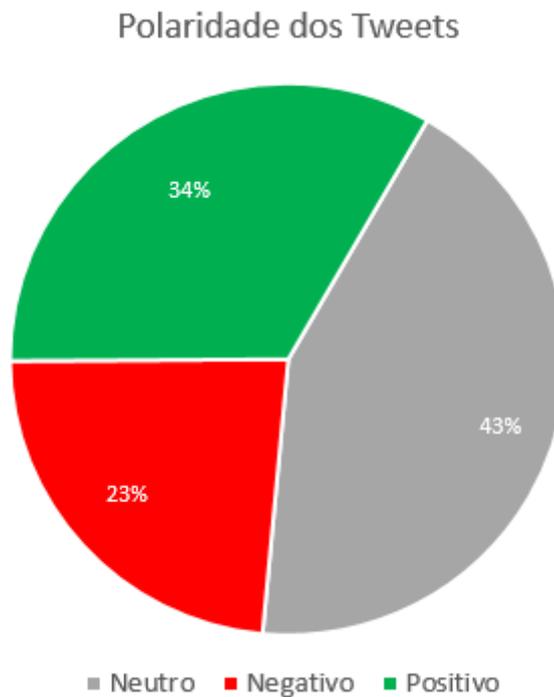
Como amostra deste projeto, foram coletados cerca de 3063 tweets durante 31 dias, devido restrições de coleta da API REST do Twitter (seção 3.2.1). Deste total, foram removidos tweets repetidos, spam, retweets e menções. Ao final deste processo de tratamento dos dados (seção 3.3), um total mínimo de 350 Tweets foram utilizados na mineração de opiniões (seção 3.4).

Neste trabalho, foi utilizado o processo que engloba as etapas de Subjetividade, Extração, Classificação e Sumarização. Ao concluir a última etapa, a ferramenta Semantria, atribui uma polaridade para cada texto, classificando em positivos, negativos e neutros.

Dos 350 tweets utilizados para análise, 117 foram classificados como positivos, 82 como negativos, e 151 como neutros, ou seja, aproximadamente 34%

dos tweets foram classificados como positivos, 23% foram classificados como negativos e os demais 43% classificados como neutros. Sendo assim, tais valores indicam que a grande maioria dos Tweets, quase metade deles foram classificados como neutros (FIGURA 16).

FIGURA 16 – CLASSIFICAÇÃO DAS POLARIDADES DOS TWEETS ANÁLISADOS



FONTE: O Autor (2017).

O levantamento do referencial teórico sobre o Bitcoin para a realização de pesquisa teste no Twitter, é apresentado no tópico 2.4. Tal referencial foi utilizado para conhecer mais sobre a tecnologia por trás do Bitcoin e também sobre a moeda.

O objetivo de criação de roteiro com os procedimentos necessários para realizar uma análise de sentimento, é apresentado no tópico 3, o qual descreve todas as etapas realizadas para uma simples análise de sentimentos no Twitter, passando pela coleta e tratamento dos dados até a análise de opinião dos mesmos.

Na sequência são apresentados os quadros contendo as facilidades e dificuldades encontradas durante a realização da análise de sentimento no Twitter. Tais quadros são fundamentais para o entendimento do que é esperado ao realizar uma análise de sentimento.

## 4.2 FACILIDADES E DIFICULDADES ENCONTRADAS NA ANÁLISE DE SENTIMENTO

Para um melhor entendimento das facilidades e dificuldades encontradas na realização da análise de sentimento, foram criados quadros contendo elementos discutidos no tópico 3. Tais quadros são divididos nas seguintes categorias: Pré-projeto, coleta e armazenamento de dados, tratamento dos dados e mineração de opiniões. O QUADRO 3 apresenta as atividades do Pré-Projeto.

QUADRO 3 - DIFICULDADES E FACILIDADES NA ETAPA DO PRÉ-PROJETO

<i>Pré-Projeto</i>			
<i>Atividade</i>	<b>Facilidades</b>	<b>Dificuldades</b>	<b>Observações</b>
<i>Escolha do tema de análise</i>	O tema não é um fator relevante para a análise de sentimento, podendo ser utilizado qualquer outros;	Necessidade de um conhecimento prévio sobre o assunto;	O Bitcoin foi escolhido como tema devido ao interesse do autor e pela falta de pesquisas nesta área.
<i>Escolha da Rede Social</i>	Criação de conta; Ampla utilização da rede social pelos usuários (mais de 300 milhões de usuários ativos); Proporciona ferramentas de interação com a rede social (APIs); Postagens com no máximo 140 caracteres;	Nenhuma dificuldade relevante encontrada	Existem outras redes sociais que podem ser utilizadas para o mesmo fim.
<i>Escolha da Ferramenta</i>	Plugin disponível para o Microsoft Excel; Análise no idioma português; Realiza a análise de opinião em todas as etapas (subjetividade, extração, classificação e sumarização)	Ferramenta é paga; Necessidade de criação da conta trial utilizando e-mail corporativo ou educacional; Limite máximo de 1000 registros analisados; Período de utilização na versão trial de 30 dias.	Existem diversos outros softwares e algoritmos de análise de sentimento que podem ser utilizadas para o mesmo fim. No entanto, este trabalho utilizou somente o Plugin Semantria;

FONTE: O Autor (2017).

Tal categoria não possui nenhuma dificuldade expressiva a ponto de prejudicar a execução do projeto, visto que para realizar uma pesquisa é necessário a definição de um tema, o que muitas vezes é atrelado a algum conhecimento prévio sobre o assunto. Em relação a escolha da rede social ou da ferramenta, existem

diversas, podendo ser necessário um estudo para identificar a que mais se adapta as necessidades do indivíduo ou organização que pretende realizar uma análise de sentimento. Este trabalho não pretende discorrer sobre as demais redes sociais e ferramentas disponíveis, de modo que apresenta um roteiro de pesquisa somente para a rede social Twitter com mineração de opinião no software Semantria.

Na sequência é mostrado o QUADRO 4, o qual discorre sobre as atividades da etapa de Coleta e armazenamento dos dados.

QUADRO 4 - DIFICULDADES E FACILIDADES NA ETAPA DE COLETA E ARMAZENAMENTO DOS DADOS

continua

<b>Coleta e Armazenamento dos Dados</b>			
<b>Atividade</b>	<b>Facilidades</b>	<b>Dificuldades</b>	<b>Observações</b>
<i>Criação de um aplicativo no Twitter</i>	Necessita somente de uma conta ativada no Twitter;	Em um primeiro momento entender se a aplicação precisava ser hospedada em algum site;	
<i>Recebimento de chaves e Token de autenticação</i>	Acessando a aplicação e navegando no menu "Keys and Access Tokens" é possível obter as chaves, somente sendo necessário gerar os Tokens;	Nenhuma dificuldade relevante encontrada	
<i>Entendimento do método GET search/tweets</i>	Existe uma ampla documentação sobre o método no site de desenvolvimento do Twitter, porém em inglês; Amplios operadores de pesquisa para utilizar na query; Existe a possibilidade de contratar um plano com acesso a todos os registros do Twitter desde 2006; Exemplos de respostas (JSON) de uma query; Parâmetros muito bem explicados e com exemplos, porém em inglês; Sistema de navegação das páginas de referência bem elaborado;	No acesso gratuito, é possível obter histórico de até 7 dias, com um limite de 100 registros coletados por vez, o que prejudica a consulta; Possibilidade de realizar 180 requisições dentro de um intervalo de 15 minutos, no entanto, a consulta traz os mesmos resultados;	

<b>Coleta e Armazenamento dos Dados</b>			
<b>Atividade</b>	<b>Facilidades</b>	<b>Dificuldades</b>	<b>Observações</b>
<i>Desenvolvimento do Script de Coleta</i>	Exemplos prontos na internet, em que só é necessário a alteração de partes do código; Possibilidade de otimizar o código de acordo com a necessidade de coleta; Existência de ferramentas gratuitas que emulam uma máquina virtual, o qual pode ser utilizado para rodar o script; Todas os parâmetros são muito bem explicados no site de referência do Twitter	Necessário conhecimento em lógica e linguagem de programação para modificar o código; Ao utilizar o plano padrão (gratuito) do Twitter, para a realização deste trabalho, foi necessário realizar a coleta cerca de 31 vezes, visto que o máximo de registros coletados por vez era somente de 100 tweets;	Este trabalho utilizou como linguagem de programação o PHP e como ferramenta de emulação de máquina virtual o Cloud 9.
<i>Definição dos atributos a serem coletados</i>	Existe uma ampla documentação sobre os atributos passíveis de coleta no site de desenvolvimento do Twitter, porém em inglês; Todos os atributos possuem exemplos e explicações, porém em inglês; Metadados podem ser escolhidos de acordo com a necessidade de coleta;	Caso não seja especificado no script de coleta, todos os atributos são coletados, trazendo inúmeras informações que não são úteis, o que torna o processamento e tratamento da resposta muito mais trabalhoso;	Para este trabalho foram coletados 7 atributos
<i>Cópia dos registros coletados</i>	Nenhuma facilidade relevante encontrada	Os resultados são retornados em uma página HTML, em que, é necessário copiar todas as informações e colar em um arquivo do bloco de notas. Feito isso, deve ser salvo no formato CSV e codificação UTF-8. Este processo precisou ser repetido 31 vezes; O processo se torna lento pela necessidade de repetição, devido a limitação de retorno da API;	
<i>Importação dos dados para o Excel</i>	Nenhuma facilidade relevante encontrada	Em alguns casos a existência do caractere ponto e vírgula (;) no atributo "Text", o que causou a quebra dos dados de forma errônea.	

FONTE: O Autor (2017).

Na etapa de coleta de dados, algumas atividades necessitam de um conhecimento mais técnico, como é o caso do desenvolvimento do script de coleta. Novamente é importante salientar que o mesmo pode ser desenvolvido em qualquer linguagem de programação, no entanto, um conhecimento no mínimo básico se torna necessário. É possível encontrar exemplos prontos na internet, no entanto, tais exemplos nem sempre atendem a necessidade total da coleta que indivíduo ou organização necessitam. Caso exista um conhecimento da linguagem de programação, é possível restringir o que realmente é importante para a coleta, diminuindo o esforço necessário no tratamento dos dados. Na FIGURA 17, é apresentado o código PHP utilizado para coletar somente os atributos que seriam necessários para a execução da mineração de opinião.

FIGURA 17 - CÓDIGO PHP UTILIZADO PARA RESTRINGIR ATRIBUTOS

```
foreach ($statuses->statuses as $stat) {  
    echo $stat->created_at . " ";  
    echo $stat->id_str . " ";  
    echo $stat->text . " ";  
    echo $stat->lang . " ";  
    echo $stat->favorite_count . " ";  
    echo $stat->retweet_count . " ";  
    echo $stat->quoted_status_id_str . " ";  
    echo "<br>";  
}
```

FONTE: O Autor (2017).

Outro fator importante para a realização de uma análise de sentimentos no Twitter é a presença de uma documentação bem detalhada a respeito de todas as funcionalidades do método e o significado de cada atributo passível de obtenção. Até o momento em que este trabalho estava sendo escrito, tal documentação estava somente disponível em Inglês. Tal documentação torna esta etapa fácil de ser realizada, no entanto podem existir alguns contratempos que demandam certo tempo para serem superados.

O QUADRO 5, apresenta as dificuldades e facilidades na etapa de tratamento dos dados.

QUADRO 5 - DIFICULDADES E FACILIDADES NA ETAPA DE TRATAMENTO DOS DADOS

<b>Tratamento dos dados</b>			
<b>Atividade</b>	<b>Facilidades</b>	<b>Dificuldades</b>	<b>Observações</b>
<i>Conversão dos campos ID para número</i>	Possibilidade de encontrar os IDs através da opção Localizar do Excel	Nenhuma dificuldade relevante encontrada	
<i>Verificação de Tweets repetidos através do metadado ID_tweet_citado</i>	Nenhuma facilidade relevante encontrada	Processo de pesquisa totalmente manual através do localizador do Excel	
<i>Validação do idioma dos Tweets coletados</i>	Possibilidade de utilizar a opção filtro do Excel	Nenhuma dificuldade relevante encontrada	
<i>Validação do atributo Texto_Tweet</i>	Nenhuma facilidade relevante encontrada	Processo manual e muito demorado, mesmo utilizando a opção filtro; Não foi encontrado nenhuma maneira automatizada para remoção de texto que possuíam atualizações do valor do Bitcoin, URLs, formas de ganhar ou minerar Bitcoin gratuitamente, e possíveis fraudes;	
<i>Verificação de tweets repetidos através do atributo Texto_Tweet</i>	Nenhuma facilidade relevante encontrada	Processo de pesquisa totalmente manual através do localizador do Excel	

FONTE: O Autor (2017).

A etapa de tratamento de dados caracterizou-se como a mais demorada e desgastante etapa deste projeto. O principal motivo é o fato da mesma ser realizada manualmente, em que não foi encontrado nenhuma ferramenta capaz de realizar de forma automatizada tal processo. Nesta etapa, era necessário remover todos os tweets repetidos (Verificação de Tweets repetidos através do metadado ID\_tweet\_citado), verificar se o idioma era português (Validação do idioma dos Tweets coletados), e validar o atributo Text\_Tweet.

Para verificar os tweets repetidos foi necessário utilizar o localizador do Microsoft Excel buscando pelo ID do tweet citado de todos os registros que possuíam esta informação. Ao encontrar um ID que possuísse um correspondente na coluna ID\_Tweet, o registro era removido. Após ser realizado a verificação dos tweets\_citados a base ficou com 2914 registros. Já na validação do atributo Idioma\_Tweet, não foi encontrado nenhum resultado diferente de “pt”.

No entanto, a verificação mais importante ocorreu na atividade de Validação do atributo Texto\_Tweet, o qual além de conter somente tweets únicos, não poderia conter spams. Foi utilizado a ferramenta filtro do Microsoft Excel, a fim de encontrar tweets com atualizações do preço do Bitcoin, URLs sem informações relevantes, como ganhar ou minerar Bitcoin gratuitamente ou possíveis fraudes. Alguns exemplos de tweets que tiveram seus registros removidos podem ser visualizados abaixo:

- 1 #Bitcoin está custando R\$9981.85 na FoxBit. Acesse <https://t.co/zbMCM99Z10> e negocie com as menores taxas do Brasil.
- O que é Bitcoin – Esquema Ponzi ou Pirâmide Financeira <https://t.co/8aOjcbU1PE> via @aulaplusbr
- BitKong - Um jogo divertido, viciante e comprovadamente justo <https://t.co/7m8dsn7stm> #Bitcoin via @bitkongme
- Descubra como Ganhar mais de 200 reais por DIA com Bitcoins <https://t.co/GLmibsGnmM> #Bitcoinmais #ganharBitcoin #Bitcoin
- Gostei de um vídeo @YouTube <https://t.co/wQdDVhwzop> Bitcoin Mining: Hardware (part1)
- O último Internet do Dinheiro! <https://t.co/DE7MoXhJQY> Agradecimentos a @BTCTN #blockchain #Bitcoin
- #Bitcoin vs #Bieber Hilarious!!! <https://t.co/VGou7ZVdcG>
- Adicionei um vídeo a uma playlist @YouTube <https://t.co/gPW7Ubasay> BITCOIN - uma formula certa para ganhar 8.000,00 reas mensal

Tais registros foram removidos por necessariamente não serem postados por usuários da rede social, e sim por bots construídos com base na API do Twitter. Além do mais, textos como os apresentados acima, não possuem uma opinião expressa, o que prejudicaria os resultados desta pesquisa, caso fossem utilizados. Após ser finalizado tal processo, 350 tweets mostraram-se aptos para a fase de mineração de opiniões.

O QUADRO 6 descreve as facilidades e dificuldades encontradas na última etapa do processo, chamada de Mineração de opiniões.

QUADRO 6 - DIFICULDADES E FACILIDADES NA ETAPA DE MINERAÇÃO DE OPINIÕES

<b>Mineração de Opiniões</b>			
<b>Atividade</b>	<b>Facilidades</b>	<b>Dificuldades</b>	<b>Observações</b>
<i>Criação de conta trial na Lexalytics</i>	Nenhuma facilidade relevante encontrada	A conta precisa de um e-mail corporativo ou educacional;	
<i>Instalação do plugin Semantria para Excel</i>	Existência de vídeos tutoriais para a instalação no próprio site da empresa;	Nenhuma dificuldade relevante encontrada	Sem necessidade de configurações avançadas, sendo somente necessário ter instalado o Microsoft Excel
<i>Autenticação via Excel no Plugin Semantria</i>	Tais informações são enviadas por e-mail e podem ser consultadas no próprio site da Lexalytics;	Nenhuma dificuldade relevante encontrada	
<i>Execução da análise de sentimento</i>	Existência de tutoriais em vídeos e texto no próprio site da empresa que ensinam a realizar a análise de sentimentos; Existência de um assistente para a execução da análise de sentimento, o qual possui alguns exemplos, tornando o processo muito intuitivo; O tempo de processamento é relativamente rápido	Limite de 1000 registros analisados e de 30 dias de utilização na versão trial, de modo que foi necessário a criação de duas contas para a execução de todos os registros;	Para este trabalho foram utilizados 350 registros
<i>Obtenção da polaridade dos tweets</i>	A ferramenta executa todas as etapas de análise de sentimento e proporciona relatórios a partir da última etapa (sumarização); Utiliza cores para diferenciar os tipos de polaridade (positivo – verde, negativo – vermelho e neutro – cinza);	Equívoco em algumas classificações;	

FONTE: O Autor (2017).

A última etapa do processo se caracteriza pela utilização da ferramenta Semantria na análise de sentimento para obter a polaridade dos tweets. Um dos pontos positivos da ferramenta é que a sua interface é intuitiva e que a mesma possui um assistente para preparar a base para ser minerada. Outro ponto que facilitou

bastante a sua utilização é o fato de ser executada dentro do Microsoft Excel, ferramenta amplamente conhecida.

A principal dificuldade na utilização desta ferramenta é o fato da mesma ser paga e a sua versão trial possuir apenas 30 dias ou 1000 análises realizadas para teste, de modo que foi necessário a criação de duas contas durante a execução total do trabalho.

A ferramenta também não se mostrou precisa na categorização de gírias ou ironias, como pode ser verificado na FIGURA 18.

FIGURA 18 - EXEMPLO DE IRONIA CLASSIFICADA INCORRETAMENTE

Source Text	Document Sentiment	Document Sentiment
Ainda bem que se tudo der errado eu tenho minha aposentadoria em Bitcoin pra daqui uns anos	-0,3	negative

FONTE: O Autor (2017).

Neste caso, o indivíduo estava comentando o fato dele estar guardando Bitcoin para compor a sua aposentadoria, e não que o fato dele guardar é algo errado. A ferramenta classificou tal sentença como negativa, no entanto, este usuário do Twitter certamente tentou expressar que o Bitcoin será o principal ativo que ele terá para a sua aposentadoria.

Outro exemplo de equívoco da ferramenta podemos observar na FIGURA 19, em que o indivíduo utiliza uma palavra comumente considerada de baixo calão, para expressar um sentimento de surpresa. Em nenhum momento a colocação de tal palavra prejudicou a mensagem que ele tentou passar, no entanto, a ferramenta classificou tal sentença como negativa.

FIGURA 19 - EXEMPLO DE EQUIVOCO EM RELAÇÃO A PALAVRAS DE BAIXO CALÃO

Source Text	Document Sentiment	Document Sentiment
Caralho, Good Wife é do século passado e já tinha episódio sobre Bitcoin!	-1,6	negative

FONTE: O Autor (2017).

A fim de apresentar alguns dos resultados obtidos com a análise de sentimento da ferramenta, o QUADRO 7 compara a análise realizada automaticamente pela mesma com uma análise realizada manualmente pelo autor da pesquisa.

QUADRO 7 - COMPARAÇÃO ENTRE ANÁLISE MANUAL E ANÁLISE AUTOMÁTICA DE TWEETS

<i>Tweet</i>	Análise Automática	Análise Manual
<i>@JorgeStolfi professor, seria sensacional uma aula sua sobre Bitcoin e a relação entre política, privacidade etc e tecnologia em MC009</i>	Positivo	Positivo
<i>@lucasmzanella Eu acho que vou comprar um 0000,1 de Bitcoin, esse sistema de segurança parece massa</i>	Positivo	Positivo
<i>Se todos os estabelecimentos de comida começarem a aceitar Bitcoin, eu vou falir.</i>	Negativo	Positivo
<i>Áustria já anunciou que as vendas realizadas em Bitcoin representavam agora 10% de suas vendas totais.</i>	Positivo	Positivo
<i>Fui em um psiquiatra para verificar a minha obsessão por Bitcoin. O resultado foi que paguei a consulta em Bitcoin e vendi mais alguns.</i>	Negativo	Positivo (Ironia)
<i>Cada vez que uso o sistema bancário meu único alívio é saber da existência do #Bitcoin e que a próxima geração rirá de lamurias como esta.</i>	Positivo	Positivo
<i>@alfredo_cunha Tá ficando bom de comprar #Bitcoin de novo.</i>	Positivo	Positivo
<i>Bitcoin é a melhor moeda que existe, estou garantindo uma boa vida com ela.</i>	Positivo	Positivo
<i>Ztenho medo do Bitcoin pq nao sei como ele funciona e o dinheiro, sabe? Os dois têm o mesmo conceito: só valem porque as pessoas confiam</i>	Neutro	Neutro
<i>Cara, se eu estivesse empregado, com certeza investiria no Bitcoin</i>	Neutro	Positivo

FONTE: O Autor (2017).

Dos 10 tweets obtidos aleatoriamente para análise, 7 deles foram classificados da mesma maneira, sendo que somente 3 deles apresentaram diferenças. Novamente em um dos casos de classificação diferente, a sentença era uma ironia, mostrando que a ferramenta possui problemas para lidar com tal tipo de colocação.

## 5 CONSIDERAÇÕES FINAIS

Atualmente, é possível utilizar a análise de sentimento como uma solução alternativa para o processo de tomada de decisão, visto que as organizações podem utilizar uma quantidade imensa de informações presentes nas redes sociais e através delas entenderem melhor seus clientes e possíveis clientes. A quantidade de registros depende da demanda de quem irá realizar a pesquisa, em que é possível contratar pacotes diretamente com as detentoras dos dados ou utilizar as opções gratuitas que as mesmas disponibilizam, possibilitando ainda assim resultados relevantes.

Este trabalho apresentou um roteiro de análise de sentimento ao analisar os comentários coletados do Twitter sobre a moeda “Bitcoin”. Estes registros foram coletados durante 31 dias através da API REST, disponibilizada pelo próprio Twitter. Ao todo foram coletados 3.063 tweets, mas foi necessário a realização de uma filtragem e tratamento dos dados, restando 350 tweets passíveis de análise. Ao finalizar a etapa de tratamento, os tweets, foram classificados em positivos, neutros ou negativos utilizando o software Semantria para a análise de sentimento.

Foi verificado na execução do roteiro de análise de sentimento que as principais dificuldades se encontram nas etapas de tratamento e mineração dos dados, em que a primeira se sucedeu de maneira totalmente manual, em que foi necessário um exaustivo esforço de verificação dos tweets, e a segunda por partes de limitação da ferramenta em classificar corretamente certos registros. O principal fator, foi a existência de gírias, ironias, abreviações ou erros de digitação que dificultaram a classificação, uma vez que o software possui a limitação de analisar somente palavras que existem em seu dicionário. Como principais facilidades é possível notar a disponibilidades de diversas APIs para a coleta de informações nas redes sociais e também diversas ferramentas de mineração de opiniões que podem ser empregados no processo.

Mesmo com algumas dificuldades e limitações encontradas no decorrer do processo, muitos tweets foram classificados corretamente e auxiliaram no entendimento das opiniões expressas sobre o Bitcoin. No entanto, vale ressaltar que é possível aprimorar de diversas maneiras o processo de mineração de opiniões, de modo que é possível criar um dicionário próprio sobre o assunto, ou agrupar dicionários existentes, permitindo que o resultado seja mais preciso.

Tal roteiro disponibilizado neste trabalho apresenta umas das maneiras mais simples de minerar opiniões, todavia, existem diversos outros métodos e algoritmos que podem ser facilmente adaptados as necessidades de cada organização ou indivíduo. Tal roteiro, mesmo sendo simples, pode ser um diferencial competitivo para as empresas aprimorarem seus produtos, serviços, processos ou até mesmo inovarem.

Por ser um campo relativamente novo de estudo, as análises obtidas nesta pesquisa, devem ser entendidas somente como indicadores. No entanto com os avanços da tecnologia, o cenário tende a evoluir ainda mais e a análise de sentimentos em redes sociais, passa a se mostrar um promissor campo de trabalho. Empresas já estão se adaptando e utilizando as mesmas como uma forma de traçar estratégias, além de utiliza-las para estabelecer um relacionamento com seus clientes.

De um modo geral, o objetivo do trabalho que buscava identificar as dificuldades e facilidades encontradas durante a execução da pesquisa, foi atingido.

## 5.1 TRABALHOS FUTUROS

Para trabalhos futuros é sugerido a utilização de outras ferramentas disponíveis no mercado, tanto para a coleta dos dados com para a mineração de opinião, criando assim um processo mais elaborado e possivelmente mais preciso e acurado.

A utilização ou criação de um dicionário próprio para o campo do conhecimento estudado, é extremamente recomendado pelo autor desde trabalho, visto que a omissão de informações por parte da ferramenta Semantria em relação a origem do seu dicionário de dados, pode prejudicar de certa forma a análise de um tema específico.

Estudos voltados para o tratamento de gírias, ironias e abreviações também são valiosos para o campo de análise de sentimento, visto que uma das dificuldades encontradas nesta pesquisa foi o da ferramenta entender e classificar corretamente tais peculiaridades da linguagem.

Para finalizar, este trabalho pode ser utilizado como uma diretriz para outras pesquisas nesta área, em que ao ler tal estudo, é possível ter uma noção das

facilidades e dificuldades esperadas na sua execução, além de ser um instrumento passível de comparação com outras técnicas ou ferramentas.

## REFERÊNCIAS

AHAMAD, S; NAIR, M; VARGHESE, B. A Survey on Crypto Currencies. **Association Of Computer Electronics And Electrical Engineers**, India, p. 42-48. abr. 2013. Disponível em: <<http://searchdl.org/public/conference/2013/AETACS/131.pdf>>. Acesso em: 21 nov. 2017.

BANCO CENTRAL DO BRASIL. **COMUNICADO Nº 25.306, DE 19 DE FEVEREIRO DE 2014**. Brasília, 2014. Disponível em: <<https://www3.bcb.gov.br/normativo/detalharNormativo.do?method=detalharNormativo&N=114009277>>. Acesso em: 21 nov. 2017

BECKER, K.; TUMITAN, D. Introdução à Mineração de Opiniões: Conceitos, Aplicações e Desafios. In: Joao Eduardo Ferreira. (Org.). **Lectures of the 28th Brazilian Symposium on Databases**. 1ed.Pernambuco: CIN - UFPE, 2013, v. , p. 27-52. Disponível em: <[http://www.inf.ufrgs.br/~kbecker/lib/exe/fetch.php?media=minicursosbbd\\_versaosubmetida.pdf](http://www.inf.ufrgs.br/~kbecker/lib/exe/fetch.php?media=minicursosbbd_versaosubmetida.pdf)>. Acesso em: 21 nov. 2017.

BERLIM, J. B. **Análise de ferramentas de mineração de opinião para aplicação em redes sociais**. 2015. 146 f. TCC (Graduação) - Curso de Gestão da Informação, Universidade Federal do Paraná. Curitiba, 2015. Disponível em: <<http://acervodigital.ufpr.br/handle/1884/41078>>. Acesso em: 28 maio 2017.

BERTAGLIA, T. F. C. **Mineração de Opinião Baseada em Extração de Aspectos**. 2015. 67 f. TCC (Graduação) - Curso de Ciência da Computação, Universidade Estadual do Oeste do Paraná, Cascavel, 2016. Disponível em: <<http://www.inf.unioeste.br/~tcc/2015/TCC - Thales Bertaglia.pdf>>. Acesso em: 21 nov. 2017.

BITCOIN.ORG. **Bitcoin**. Disponível em: <<https://bitcoin.org/en/>>. Acesso em: 21 nov. 2017.

BITCOIN BRASIL. **Bitcoin Brasil**. Disponível em: <<https://www.bitcoinbrasil.com.br>>. Acesso em: 21 nov. 2017.

BLOCKEXPLORER. **Bitcoin Block Explorer**. Disponível em: <<https://blockexplorer.com>>. Acesso em: 21 nov. 2017.

BRASIL. Câmara dos Deputados. Projeto de Lei N.º 2.303, de 2015. Autor: Aureo Lídio Moreira Ribeiro. **Câmara dos Deputados**, Brasília, DF, 08 jul. 2015. Disponível em: <<http://www.camara.gov.br/sileg/integras/1361896.pdf>>. Acesso em: 21 nov. 2017.

BRITO, J.; CASTILLO, A. **Bitcoin: a primer for policymakers**. Arlington: Mercatus Center, 2013. Disponível em: <[https://www.mercatus.org/system/files/Brito\\_BitcoinPrimer.pdf](https://www.mercatus.org/system/files/Brito_BitcoinPrimer.pdf)>. Acesso em: 21 nov. 2017.

CANALTECH. **O que é API?**. Disponível em: <<https://canaltech.com.br/o-que-e/software/o-que-e-api/>>. Acesso em: 21 nov. 2017.

CLOUD9. **Cloud9**. Disponível em: <<https://c9.io>>. Acesso em: 21 nov. 2017.

COINMAP. **Coinmap 2.0**. Disponível em: <<https://coinmap.org/welcome/>>. Acesso em: 21 nov. 2017.

ELSEVIER. **Scopus**. Disponível em: <<https://www.elsevier.com/solutions/scopus>>. Acesso em: 21 nov. 2017.

FERREIRA, E de B. A. **Análise De Sentimento Em Redes Sociais Utilizando Influência Das Palavras**. 2010. 69 f. TCC (Graduação) - Curso de Ciência da Computação, Universidade Federal de Pernambuco, Recife, 2010. Disponível em: <<http://www.cin.ufpe.br/~tg/2010-2/ebaf.pdf>>. Acesso em: 21 nov. 2017.

FOLHA DE SÃO PAULO. **Brasileiro tem que declarar Bitcoin; IR pode ser cobrado**. 2014. Disponível em: <<http://www1.folha.uol.com.br/mercado/2014/04/1436809-brasileiro-tem-que-declarar-bitcoin-ir-pode-ser-cobrado.shtml>>. Acesso em: 21 nov. 2017.

G1. **1ª loja física do Brasil a vender Bitcoin quer levar moeda virtual para a rua**. 2014. Disponível em: <<http://g1.globo.com/tecnologia/noticia/2014/06/1-loja-fisica-do-brasil-vender-bitcoin-quer-levar-moeda-virtual-para-rua.html>>. Acesso em: 21 nov. 2017.

G1. **Facebook atinge os 2 bilhões de usuários**. 2017. Disponível em: <<https://g1.globo.com/tecnologia/noticia/facebook-atinge-os-2-bilhoes-de-usuarios.ghtml>>. Acesso em: 21 nov. 2017.

GABRIEL, M. **Marketing na Era Digital: Conceitos, Plataformas e Estratégias**. São Paulo: Novatec, 2010. Disponível em: <<https://kamilamendonca.files.wordpress.com/2015/09/marketing-na-era-digital.pdf>>. Acesso em: 21 nov. 2017.

GARCIA, R. de S. **Moedas virtuais são moedas?: um estudo de caso para o Bitcoin e o Litecoin**. 2014. 36 f. TCC (Graduação) - Curso de Ciências Econômicas, Universidade Estadual de Campinas, Campinas, 2014. Disponível em: <[www.bibliotecadigital.unicamp.br/document/?down=000944475](http://www.bibliotecadigital.unicamp.br/document/?down=000944475)>. Acesso em: 21 nov. 2017.

HISTORY OF BITCOIN. **History of Bitcoin: The world's first decentralized currency**. Disponível em: <<http://historyofBitcoin.org>>. Acesso em: 21 nov. 2017.

LEXALYTICS: **Semantria for Excel**. Disponível em: <<https://www.lexalytics.com/semantria/excel>>. Acesso em: 21 nov. 2017.

LIU, Bing. **Sentiment Analysis and Opinion Mining**. [S.l.]: Morgan & Claypool Publishers, 2012. 168 p. Disponível em:

<<https://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf>>. Acesso em: 21 nov. 2017.

MARTELETO, R. M. Análise de Redes Sociais: Aplicação nos Estudos de Transferência da Informação. **Ciência da Informação**. Brasília, v. 30, n. 1, p. 71-81. jan. 2001. Disponível em: <<http://www.scielo.br/pdf/ci/v30n1/a09v30n1>>. Acesso em: 21 nov. 2017.

NAKAMOTO, S. **Bitcoin**: a Peer-to-Peer Electronic Cash System, 2008. Disponível em: <<https://Bitcoin.org/Bitcoin.pdf>>. Acesso em: 21 nov. 2017.

NOGUEIRA, E. D. A. **Análise de Brand Equity sob a Perspectiva do Consumidor nas Mídias Sociais por Meio da Mineração de Opinião e Análise de Redes Sociais**. 2015. 236 f. Dissertação (Mestrado) - Curso de Programa de Pós-Graduação em Ciência, Gestão e Tecnologia da Informação, Universidade Federal do Paraná. Curitiba, 2015. Disponível em: <<http://acervodigital.ufpr.br/handle/1884/41257>>. Acesso em: 21 nov. 2017.

PANG, B; LEE, L. Opinion mining and sentiment analysis. **Foundations And Trends In Information Retrieval**. Boston, USA, v. 2, n. 1-2, p. 1-135. jul. 2008. Disponível em: <<http://www.cs.cornell.edu/home/llee/omsa/omsa.pdf>>. Acesso em: 21 nov. 2017.

PANG, B.; LEE, L.; VAITHYANATHAN, S. Thumbs up? Sentiment Classification using Machine Learning Techniques. **Proceedings Of The 2002 Conference On Empirical Methods In Natural Language Processing**. Philadelphia, Usa, p. 79-86. jul. 2002. Disponível em: <<http://dl.acm.org/citation.cfm?id=1118704>>. Acesso em: 21 nov. 2017.

PINTO, A. M. G.; JUNQUEIRA, L. A. P. Relações de poder em uma rede do terceiro setor: um estudo de caso. **Revista de Administração Pública**, Rio de Janeiro, v. 43, n. 5, p. 1091-1116, 2009. Disponível em: <<http://www.scielo.br/pdf/rap/v43n5/v43n5a06.pdf>>. Acesso em: 21 nov. 2017.

RECUERO, R. **Redes Sociais Na Internet**. Porto Alegre: Editora Sulina, 2009. Disponível em: <<http://www.ichca.ufal.br/graduacao/biblioteconomia/v1/wp-content/uploads/redessociaisnainternetrecuero.pdf>>. Acesso em: 21 nov. 2017.

SILVA, D. E da. Aspectos de segurança na rede Bitcoin. In: ALMEIDA, F. A. S. de. et al. **Coletânea Luso-Brasileira/ Gestão da Informação, Cooperação em Redes e Competitividade**. Porto: Universidade do Porto, 2014. Cap. 8. p. 221-249. Disponível em: <[http://www.cdn.ueg.br/source/observatorio\\_inhumas/conteudoN/3322/CAP\\_8\\_CR\\_ASPECTOS\\_DE\\_SEGURANCA\\_NA\\_REDE\\_BITCOIN.pdf.pdf](http://www.cdn.ueg.br/source/observatorio_inhumas/conteudoN/3322/CAP_8_CR_ASPECTOS_DE_SEGURANCA_NA_REDE_BITCOIN.pdf.pdf)>. Acesso em: 21 nov. 2017.

SILVA, N. G. R. da. **BestChoice**: classificação de sentimento em ferramentas de expressão de opinião. 2010. 54 f. TCC (Graduação) - Curso de Ciência da Computação, Universidade Federal de Pernambuco, Recife, 2010. Disponível em: <<http://www.cin.ufpe.br/~tg/2010-2/ngrs.pdf>>. Acesso em: 21 nov. 2017.

SOCIALMENTION. **Socialmention**: Real-time social media search and analysis. Disponível em: <<http://www.socialmention.com>>. Acesso em: 21 nov. 2017.

SOUZA, T. B. de; CATARINO, M. E.; SANTOS, P. C. dos. Metadados: Catalogando dados na internet. **Transinformação**, Campinas, v. 9, n. 2, p.93-105, abr. 1997. Disponível em: <<http://periodicos.puc-campinas.edu.br/seer/index.php/transinfo/article/view/1586/1558>>. Acesso em: 23 nov. 2017.

STATISTIC BRAIN. **Twitter Statistics**. Disponível em: <<http://www.statisticbrain.com/twitter-statistics/>>. Acesso em: 21 nov. 2017.

TECMUNDO. **A história do Twitter**. 2010. Disponível em: <<https://www.tecmundo.com.br/rede-social/3667-a-historia-do-twitter.htm>>. Acesso em: 21 nov. 2017.

TECNISA. **Bitcoins na Tecnisa**. Disponível em: <<https://www.tecnisa.com.br/lp/bitcoins>>. Acesso em: 21 nov. 2017.

TWITTER DEVELOPERS. **Twitter Developer Documentation**. Disponível em: <<https://dev.twitter.com/docs>>. Acesso em: 21 nov. 2017.

TWITTER. **Twitter**. Disponível em: <<https://twitter.com>>. Acesso em: 21 nov. 2017.

ULRICH, F. **Bitcoin A Moeda Na Era Digital**. São Paulo: Instituto Ludwig von Mises Brasil, 2014. Disponível em: <[http://www.informatrader.com.br/datafiles/conteudo\\_downloads/9/Fernando-Ulrich-Bitcoin.pdf](http://www.informatrader.com.br/datafiles/conteudo_downloads/9/Fernando-Ulrich-Bitcoin.pdf)>. Acesso em: 21 nov. 2017.

VEJA. **Caixa eletrônico de bitcoins movimenta R\$ 180 mil em dois meses**. Disponível em: <<http://veja.abril.com.br/tecnologia/caixa-eletronico-de-bitcoins-movimenta-r-180-mil-em-dois-meses/>>. Acesso em: 21 nov. 2017.

VERAS, M de O. **As Redes Sociais Como Instrumento De Ensino-Aprendizagem Através Da Educomunicação: Análise A Partir Do Facebook**. 2013. 112 f. TCC (Graduação) - Curso de Comunicação Social, Faculdades Nordeste, Fortaleza, 2013. Disponível em: <<https://pt.slideshare.net/marciodveras/monografia-marcio-verso-pdf>>. Acesso em: 21 nov. 2017.

WEB OF SCIENCE. **Web of Science**. Disponível em: <<https://webofknowledge.com>>. Acesso em: 21 nov. 2017.

## APÊNDICE A – RETORNO DE UMA CONSULTA NO FORMATO JSON

```
[1] => stdClass Object
(
    [created_at] => Wed May 31 23:52:51 +0000 2017
    [id] => 870065570334560257
    [id_str] => 870065570334560257
    [text] => Gostei de um vídeo @YouTube https://t.co/H1XsWHoueK] Como
Ganhar 180 Mil Satoshis / Dia [PT-BR]
    [truncated] =>
    [entities] => stdClass Object
    (
        [hashtags] => Array
        (
        )

        [symbols] => Array
        (
        )

        [user_mentions] => Array
        (
            [0] => stdClass Object
            (
                [screen_name] => YouTube
                [name] => YouTube
                [id] => 10228272
                [id_str] => 10228272
                [indices] => Array
                (
                    [0] => 19
                    [1] => 27
                )
            )
        )

    )

    [urls] => Array
```

```

(
  [0] => stdClass Object
  (
    [url] => https://t.co/H1XsWHoueK
    [expanded_url] => http://youtu.be/TojYDjjiurA?a[Bitcoin
    [display_url] =>youtu.be/TojYDjjiurA?a[...]
    [indices] => Array
      (
        [0] => 28
        [1] => 51
      )
    )
  )
)

[metadata] => stdClass Object
(
  [iso_language_code] => pt
  [result_type] => recent
)

[source] => <a href="http://www.google.com/" rel="nofollow">Google</a>
[in_reply_to_status_id] =>
[in_reply_to_status_id_str] =>
[in_reply_to_user_id] =>
[in_reply_to_user_id_str] =>
[in_reply_to_screen_name] =>
[user] => stdClass Object
(
  [id] => 835091086943076352
  [id_str] => 835091086943076352
  [name] => Adriel Beep
  [screen_name] => AdrielGazz
  [location] =>
  [description] =>
  [url] =>

```

[entities] => stdClass Object

```
(
  [description] => stdClass Object
  (
    [urls] => Array
    (
    )
  )
)
```

[protected] =>

[followers\_count] => 3

[friends\_count] => 7

[listed\_count] => 0

[created\_at] => Fri Feb 24 11:36:45 +0000 2017

[favourites\_count] => 0

[utc\_offset] => -25200

[time\_zone] => Pacific Time (US & Canada)

[geo\_enabled] =>

[verified] =>

[statuses\_count] => 187

[lang] => pt

[contributors\_enabled] =>

[is\_translator] =>

[is\_translation\_enabled] =>

[profile\_background\_color] => F5F8FA

[profile\_background\_image\_url] =>

[profile\_background\_image\_url\_https] =>

[profile\_background\_tile] =>

[profile\_image\_url]

=>

[http://abs.twimg.com/sticky/default\\_profile\\_images/default\\_profile\\_normal.png](http://abs.twimg.com/sticky/default_profile_images/default_profile_normal.png)

[profile\_image\_url\_https]

=>

[https://abs.twimg.com/sticky/default\\_profile\\_images/default\\_profile\\_normal.png](https://abs.twimg.com/sticky/default_profile_images/default_profile_normal.png)

[profile\_link\_color] => 1DA1F2

[profile\_sidebar\_border\_color] => C0DEED

[profile\_sidebar\_fill\_color] => DDEEF6

[profile\_text\_color] => 333333

```
[profile_use_background_image] => 1  
[has_extended_profile] =>  
[default_profile] => 1  
[default_profile_image] => 1  
[following] =>  
[follow_request_sent] =>  
[notifications] =>  
[translator_type] => none  
)
```

```
[geo] =>  
[coordinates] =>  
[place] =>  
[contributors] =>  
[is_quote_status] =>  
[retweet_count] => 0  
[favorite_count] => 0  
[favorited] =>  
[retweeted] =>  
[possibly_sensitive] =>  
[lang] => pt  
)
```

## APÊNDICE B – SCRIPT PARA COLETA DE TWEETS

```

<!DOCTYPE html>
<html >
<header>
  <meta charset="utf-8" />
</header>
<body>

<?php

// Keys and tokens
$consumer_key = 'Inserir a chave de sua aplicação disponibilizada pelo Twitter';
$consumer_secret = 'Inserir a chave de sua aplicação disponibilizada pelo Twitter';
$access_token = 'Inserir o Token de sua aplicação disponibilizada pelo Twitter';
$access_token_secret = 'Inserir o Token de sua aplicação disponibilizada pelo Twitter';

//Include library OAuth
require "twitteroauth/autoload.php";
use Abraham\TwitterOAuth\TwitterOAuth;

//Connect to API
$connection = new TwitterOAuth($consumer_key, $consumer_secret, $access_token,
$access_token_secret);
$content = $connection->get("account/verify_credentials");

//get_limits
//$test = $connection->get("application/rate_limit_status");
//print_r($test);

//get
$statuses = $connection->get("search/tweets", ["q" => "Bitcoin", "count" => 100, "until" => "
2017-06-04", "result_type" => "recent", "lang" => "pt"]);

foreach ($statuses->statuses as $stat) {
  echo $stat->created_at . ";";
  echo $stat->id_str . ";";
  echo $stat->text . ";";
  echo $stat->lang . ";";
}

```

```
    echo $stat->favorite_count . " ";  
    echo $stat->retweet_count . " ";  
    echo $stat->quoted_status_id_str . " ";  
    echo "<br>";  
  }  
?>  
</body>  
</html>
```