Universidade Estadual de Londrina

GABRIEL MARTINS MAIA

DEMOCRACIA ELETRÔNICA: UM ESTUDO DE SUAS PRINCIPAIS TECNOLOGIAS

LONDRINA - PR

2013

GABRIEL MARTINS MAIA

DEMOCRACIA ELETRÔNICA: UM ESTUDO DE SUAS PRINCIPAIS TECNOLOGIAS

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Ciência da Computação da Universidade Estadual de Londrina para obtenção do título de Bacharel em Ciência da Computação

Orientador: Prof. Dr. Bruno Bogaz Zarpelão.

LONDRINA - PR

2013

GABRIEL MARTINS MAIA

DEMOCRACIA ELETRÔNICA: UM ESTUDO DE SUAS PRINCIPAIS TECNOLOGIAS

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Ciência da Computação da Universidade Estadual de Londrina para obtenção do título de Bacharel em Ciência da Computação

BANCA EXAMINADORA

_____

Prof. Dr. Bruno Bogaz Zarpelão
Universidade Estadual de Londrina

_____

Prof. Dr. Mario Lemes Proença Jr.
Universidade Estadual de Londrina

_____

Prof. Dr. Sylvio Barbon Jr.
Universidade Estadual de Londrina

Londrina, _____de _____de _____.

Aos meus pais Angelo Antonio Maia e Rosana Martins Maia.

A meus amigos de Cândido Mota.

A meu irmão Guilherme Martins Maia e todos familiares.

Ao meu orientador e todos professores.

AGRADECIMENTOS

Queria agradecer a todos que sempre acreditaram em mim, principalmente minha família e amigos. Também gostaria de agradecer ao meu orientador prof. Dr. Bruno Bogaz Zarpelão e ao professor Sylvio Barbon Jr. que deram total apoio durante todo ano.

“Vamos agradecer aos idiotas. Não fosse por eles não faríamos tanto sucesso.”

(Mark Twain)

MAIA, Gabriel Martins. **Democracia Eletrônica: Um Estudo de suas Principais Tecnologias**. 2013. 50 p. Trabalho de Conclusão de Curso (Graduação em Ciência da Computação) – Universidade Estadual de Londrina, Londrina, 2013.

RESUMO

Com o desenvolvimento rápido das Tecnologias de Informação e Comunicação, o termo democracia eletrônica (em inglês, electronic democracy ou e-democracy) começou a ganhar novas proporções. Trata-se de uma nova forma de fazer com que o cidadão comum participe em discussões e interações com os poderes políticos, fazendo chegar a sua voz e suas opiniões. Com o surgimento de novos meios e canais de participação democrática, este trabalho pretende analisar a literatura recente sobre Web 2.0, mídias sociais e redes sociais aplicados ao desenvolvimento de soluções de democracia eletrônica. Além disso, foram realizados alguns testes com a aplicação do algoritmo TF-IDF (Term Frequency–Inverse Document Fequency) para mineração de textos publicados sobre alguns prefeitos e governadores brasileiros no Twitter.

Palavras-chave: democracia eletrônica. Web 2.0. redes sociais. mineração de texto.

MAIA, Gabriel Martins. Electronic Democracy: a Study of Main Technologies. 2013. 50 p. Final Project (Undergraduation). Bachelor of Science in Computer Science – State University of Londrina, 2013..

ABSTRACT

With the rapid development of Information and Communication Technology, the term e-democracy began to gain new proportions. It is a new way to make ordinary citizens to participate in discussions and interactions with the political powers, to bring his voice and opinions. With the emergence of new media and channels of democratic participation, this work examines recent research on Web 2.0, social media and social networking solutions applied to the development of e-democracy. Moreover, this work presents some results on the application of TFIDF text mining algorithm on posts in Twitter related to Brazilian mayors and governors.

Keywords: e-democracy. Web 2.0. social networks. text mining.

LISTA DE ILUSTRAÇÕES

Figura 2.1 - resultado da obra vencedora em duas regiões [22] 27
Figura 3.1 - Etapas da mineração de texto 34
Figura 3.2 - Cenário da Aplicação 39
Figura 3.3 - Modelo de entidade relacionamento do banco usado 39

LISTA DE TABELAS

Tabela 2.1 – Modelos de democracia eletrônica [9] 16
Tabela 2.2 – Participação do Cidadão no Orçamento Participativo [21].....28
Tabela 3.1 – Prefeito de São Paulo 41
Tabela 3.2 – Prefeito do Rio de Janeiro 42
Tabela 3.3 – Governador do Rio de Janeiro 42
Tabela 3.4 – Prefeito de Curitiba 43
Tabela 3.5 – Governador do Paraná.....43

LISTA DE ABREVIATURAS E SIGLAS

PLN	Processamento de Linguagem Natural
TF-IDF	Term Frequency-Inverse Document Frequency
API	Application Programming Interface
REST	Representational State Transfer

SUMÁRIO

1 INTRODUÇÃO 13

2 FUNDAMENTAÇÃO TEÓRICA 15

2.1 DEMOCRACIA ELETRÔNICA 15

2.1.1 Modelos de Democracia Eletrônica 16

2.1.2 Setores da Democracia Eletrônica 18

2.1.3 Desafios da Democracia Eletrônica 21

2.2 TECNOLOGIAS UTILIZADAS NA DEMOCRACIA ELETRÔNICA 21

2.2.1 Web 2.0 22

2.2.2 Redes Sociais 28

2.2.3 Processamento de Linguagem Natural (PLN) 29

3 DEMOCRACIA ELETRÔNICA E MINERAÇÃO DE TEXTO NO TWITTER 32

3.1 TWITTER 32

3.2 MINERAÇÃO DE TEXTO 33

3.2.1 Etapas da Mineração de Texto 34

3.2.2 TF-IDF 36

3.3 LEVANTAMENTO DE PALAVRAS MAIS SIGNIFICATIVAS EM *TWEETS* DIRECIONADOS A GOVERNANTES 37

4 CONCLUSÃO 45

5 REFERÊNCIAS BIBLIOGRÁFICAS 46

1 INTRODUÇÃO

Com a ajuda da Tecnologia de Informação e Comunicação (TIC), especialmente a Internet, um governo melhor é visto como sendo mais sensível às necessidades de cidadãos e empresas [1]. Essas tecnologias mudaram a abordagem sobre gestão pública, configurando uma realidade na qual cidadãos, empresas e demais organizações querem ter acesso cada vez mais rápido e fácil ao governo, na expectativa de que suas necessidades sejam atendidas pelos programas governamentais [2].

A interdependência da tecnologia e da sociedade é inquestionável e tem influenciado a pesquisa em ciências exatas, bem como em ciências sociais por um longo tempo. Em especial, tecnologias para apoiar a participação do cidadão na gestão pública e democracia eletrônica podem potencialmente influenciar todo um sistema político [3].

A democracia eletrônica reúne uma série de processos e ferramentas utilizados pelos governantes para facilitar uma maior participação da população no governo através de meios digitais ou eletrônicos. Estas iniciativas podem incluir e-fóruns, reuniões, e-consultas, referendos e e-voto além de outras formas de e-participação [4].

Este trabalho de conclusão de curso tem dois objetivos principais. O primeiro é realizar um levantamento dos conceitos e estudos que envolvem a democracia eletrônica. O segundo é discutir a aplicação de mineração de texto em redes sociais para democracia eletrônica. Será feito um experimento com a aplicação do algoritmo de mineração de texto TF-IDF (*Term Frequency-Inverse Document Frequency*) em *posts* do Twitter. Neste experimento, serão coletados *posts* no microblog Twitter que estejam relacionados a prefeitos e governadores brasileiros. Estes posts serão analisados com o TF-IDF, de forma a descobrirmos quais foram as palavras mais relevantes e qual é a relação delas com a realidade.

A organização deste trabalho é realizada da seguinte maneira. Na seção 2, encontra-se a fundamentação teórica e estudo do estado da arte dos principais assuntos e técnicas referentes a democracia eletrônica.

A seção 3 apresenta um experimento no qual o algoritmo de mineração de texto TF-IDF foi aplicado a *posts* destinados a prefeitos e governadores no Twitter. O objetivo da seção é mostrar que o Twitter pode refletir os assuntos mais tratados no cotidiano,

trazendo informações importantes ao governante e facilitando a participação popular no governo. Finalmente, no capítulo 4 serão apresentadas as conclusões deste trabalho e sugestões para trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Esta seção mostrará o que vem a ser democracia eletrônica, como também serão abordados as principais tecnologias e trabalhos relacionados a democracia eletrônica, mostrando suas principais características e como elas podem ajudar a relação de participação da população nas decisões governamentais.

2.1 DEMOCRACIA ELETRÔNICA

A democracia eletrônica é a convergência do processo democrático tradicional com processos e Tecnologias da Informação e Comunicação (TIC). O termo democracia eletrônica refere-se ao modo como as TIC, principalmente a Internet, podem ser usadas para melhorar os nossos processos democráticos e fornecer maiores oportunidades para indivíduos e comunidades para interagir com o governo. A democracia eletrônica compreende uma gama de atividades baseadas nas TIC, principalmente a Internet, que visam fortalecer a democracia e instituições, incluindo agências governamentais [1]. Para outros autores, democracia eletrônica também utiliza as TIC, mas tem como objetivo proporcionar o acesso dos cidadãos à informação e conhecimento sobre o processo político e as escolhas [3]. A TIC aplicada à governança urbana oferece um enorme potencial para alcançar simultaneamente uma gestão eficaz da cidade e envolver os cidadãos na definição de seus futuros.

De acordo com [5], democracia eletrônica é tudo que os governos fazem para facilitar uma maior participação popular no governo através de meios digitais ou eletrônicos, e também um meio para maior divulgação de informações políticas e para melhorar a comunicação e participação.

Por meio da utilização de diferentes formas de TIC, como a Internet e dispositivos móveis, há a oportunidade não só para realizar um trabalho mais eficaz, mas também entrar em contato com aqueles que normalmente não têm nenhuma participação nas questões políticas. Numa perspectiva *bottom-up*, os cidadãos e organizações podem usar as TIC como recurso para que sua voz seja ouvida, partidos políticos as utilizam para a

campanha, enquanto as agências públicas utilizam para melhorar a qualidade dos serviços que elas entregam aos cidadãos. [6]

A participação do público e a permissão para que um público mais amplo contribua para o debate político baseado em computação social (redes sociais) e tecnologias móveis estão transformando a maneira como as pessoas usam a Internet para comunicar e interagir. A Tecnologia da Informação e Comunicação (TIC) é vista como facilitador da participação popular (e-participation) e da democracia (e-democracy) [7].

Como o processo de governo e política se torna mais transparente, participativo e online, há uma grande oportunidade para a adoção de ferramentas baseadas em inteligência artificial, Web 2.0, sistemas avançados de pesquisas para e-governo e política. [8].

2.1.1 Modelos de Democracia Eletrônica

Os modelos de democracia eletrônica são estruturas que se relacionam com o uso da tecnologia para as diversas formas de organizações políticas, enfatizando, principalmente, o impacto das TIC nos processos de tomada de decisão pública. [9]

Os quatro modelos de democracia eletrônica são classificados com base em duas características fundamentais definidas em relação a qualquer processo democrático: a inclusão das pessoas (cidadão) em decisões e controle da agenda. Controle da agenda define quem decide e o que deve ser decidido, o que dá o direito aos cidadãos de levantar questões e participar ativamente nos processos de tomada de decisão. [9]

Tabela 2.1- Modelos de democracia eletrônica [9]

	Democracia eletrônica partidária	Democracia eletrônica direta
Cidadãos Definem Agenda	Cidadãos expressam opiniões e críticas sobre as estruturas de poder existentes. Cidadãos definem a agenda de discussões públicas, mas não para a tomada de decisões.	Os cidadãos participam diretamente nos processos de decisão. Os cidadãos estão online, afetando as decisões a serem tomadas. Cidadãos definem a agenda, tanto para a discussão pública e tomada de decisão.

	Democracia eletrônica liberal	Democracia eletrônica deliberativa
Políticos funcionários definem Agenda	Governo atende cidadãos que participam nas eleições e debates relacionados, para poder informar e ser informado por parte dos cidadãos, sem uma clara conexão com o processo de tomada de decisão.	Projetos de democracia eletrônica são utilizados para fins específicos que envolvam os cidadãos nos processos de tomada de decisões públicas. Os cidadãos têm uma boa razão para esperar que suas vozes sobre um determinado assunto sejam ouvidas.

A democracia eletrônica liberal opera seguindo os princípios do liberalismo, no qual se caracteriza as eleições partidárias, assim as agências governamentais e a elite política definem a agenda para os procedimentos para tomada de decisão. A democracia eletrônica liberal é baseada em um governo representativo, no qual os cidadãos formam o eleitorado, participam do debate e elegem representantes, mas não participam do processo de tomada de decisão [9] [10].

A democracia eletrônica deliberativa conecta os cidadãos de forma mais explícita e direta aos processos de tomada de decisão, desta forma a votação não é a principal forma de se legitimar. Deste modo, os cidadãos junto com os políticos podem dialogar em relação as ideias que levam à formação da política pública. Esta é uma forma de democracia representativa, onde a entrada transparente e cooperação entre os cidadãos e políticos constituem a legalização do poder [9] [11].

Democracia eletrônica partidária é independente ou em oposição às estruturas de poder existentes, cidadãos expressam opiniões e críticas sobre as estruturas de poder existentes, deste modo eles participam na definição da agenda para discussões públicas, mas não para a tomada de decisões [9] [11].

O modelo de democracia eletrônica direta, ao invés de, uma democracia representativa em que as pessoas votam em representantes que, em seguida, decidem iniciativas políticas, representa uma alternativa radical no qual cidadãos participam diretamente nos processos de decisão, definem a agenda, tanto para a discussão pública quanto para tomada de decisão. Com isso, a Internet não apenas faz parte do processo, mas se

torna essencial para este modelo de democracia, onde as questões são decididas pelo povo [9] [12].

2.1.2 Setores da Democracia Eletrônica

O processo de democracia eletrônica inclui *e-parliament, e-legislation, e-justice, e-mediation, e-environment, e-voting, e-consultation, e-participation, e-initiatives, e-petitioning, e-campaigning e-polling/e-surveying,* cada um destes setores são explicados [4] [6] a seguir.

E-voting **(voto eletrônico)***:* é uma votação ou referendo que envolve o uso de meios eletrônicos, possibilitando maior número de envolvidos.

E-participation **(participação eletrônica)***:* os cidadãos têm mais poder sobre a formulação de políticas e processos de tomada de decisão, criando meios para que os cidadãos possam participar do processo.

E-petitioning **(petição eletrônica)***:* petição eletrônica ou protesto que é entregue para uma instituição democrática, onde cidadãos assinam uma petição e, possivelmente, se envolvem em uma discussão sobre o assunto.

E-polling/e-surveying **(pesquisa eletrônica)**: pesquisas e coleta de informações e opiniões dos cidadãos por meio eletrônico, geralmente com respostas de múltiplas escolhas.

E-consultation **(consultas eletrônicas)***:* é uma maneira de coletar as opiniões do cidadão ou do público em geral sobre a questão política específica, podendo ou não adotar o que foi coletado.

E-legislation **(legislação/lei eletrônica)***:* é o uso das TIC para comentar, consultar, estruturar, formatar, apresentar, alterar e votar as leis que passam por assembleias eleitas, aumentando a transparência e participação da população.

E-environment **(leis ambientais eletrônicas)***:* é a utilização das TIC para fins de avaliação ambiental e proteção e o uso sustentável dos recursos naturais, aumentando a transparência e participação com questões ambientais.

E-campaigning **(campanha eletrônica)***:* utiliza-se meio eletrônico, incentivando as pessoas a interagir uma com as outras, a fim de mobilizar os indivíduos em campanhas eleitorais e, influenciar a formação e implementação de políticas públicas.

E-parliament **(parlamento/assembleia legislativa eletrônicos)***:* membros e pessoal político e administrativo de assembleias eleitas utilizam a TIC para a condução das suas atribuições, em especial, para fins de participação ativa dos cidadãos.

E-justice **(justiça eletrônica)***:* é a utilização das TIC na realização de justiça por todos do sistema judiciário, a fim de melhorar a eficiência e a qualidade dos serviços públicos, em particular para indivíduos e empresas, que inclui a comunicação eletrônica e troca de dados, bem como o acesso à informação judicial.

E-initiatives **(iniciativas eletrônicas)***:* permitir que os cidadãos façam propostas políticas por meio das TIC, envolvendo-os na definição da agenda política.

E-mediation **(acordos/mediação eletrônicos)***:* é o uso das TIC para encontrar os meios para resolução de disputas (acordos) sem a presença física das partes.

Um grande exemplo de como as TIC podem ajudar o cidadão participar mais dos processos políticos, uma ferramenta conhecida como *Smartvote*, tem sida usada para referendos e eleições.

No decorrer desta última década, o número de pessoas que passaram a ter acesso à Internet cresceu de maneira clara, e isso deu grandes esperanças para melhoria na qualidade democrática.

O voto eletrônico deve levar cada vez mais eleitores a participar do processo eletivo, pois o acesso à informação fica mais fácil em sites políticos que podem fornecer aos cidadãos uma ampla variedade de informações sobre os problemas políticos, os partidos, os candidatos e fóruns on-line, blogs e chats devem promover discussões entre os cidadãos, bem como entre os cidadãos e a elite política. Contudo, a tecnologia da Internet não oferece só uma ampla variedade de possibilidades para difundir informação, mas também abre completamente novas formas de processos de tomada de decisão [13].

Deste modo, Fivaz e Schwarz [13] propuseram uma ferramenta chamada *Smartvote* para o auxílio na escolha de voto. O site *Smartvote* consiste em três elementos principais: o módulo *Smartvote* com o sistema de emissão de correspondência, um banco de

dados completo de informações de todos os candidatos que concorrem a um cargo, e um módulo com análise elaborada em ferramentas para visualização de posições políticas.

Um exemplo do funcionamento da ferramenta *Smartvote* será feito nos parágrafos abaixo:

Numa primeira fase, cerca de três meses antes do dia da eleição, todos os candidatos recebem um questionário, com cerca de 70 questões sobre várias posições políticas, para assim estabelecer seu perfil.

Cerca de seis semanas antes de o período eleitoral começar, a segunda fase tem início. Agora o website do *Smartvote* está aberto para os eleitores, que precisam seguir três etapas para concluir a recomendação do voto.

A primeira etapa consiste em os eleitores especificarem o seu perfil político. Logo, eles são convidados a responder ao mesmo questionário que os candidatos já responderam. As perguntas são de fácil entendimento e ainda possuem um apoio adicional com informações e esclarecimentos, incluindo os pontos positivos e negativos de cada questão. Em contraste com os candidatos, os eleitores têm também a opção "nenhuma resposta", e podem pesar as respostas de acordo com a relevância.

Depois de concluída a primeira etapa, os eleitores têm de personalizar a sua recomendação de voto, que é a segunda etapa. Depois de definir o seu perfil político, é preciso especificar em qual distrito eleitoral vota e qual campanha, disputa querem receber uma recomendação de voto. Também podem escolher uma recomendação de voto para partidos ou em candidatos individuais.

Com o resultado deste questionário, um perfil político sobre o usuário é feito. Com estas informações o *Smartvote* compara as respostas dos eleitores com as respostas de um candidato, incluindo o peso que cada eleitor deu às perguntas. Quanto maior for à congruência das respostas entre o eleitor e o candidato, maior é a sua ‘afinidade’ com o candidato. O cálculo da recomendação de voto individual de cada eleitor é feita através do escalonamento multidimensional.

Este processo é repetido com todas as perguntas e para cada candidato do selecionado eleitorado. Em segundo lugar, *Smartvote* gera a recomendação de voto em forma de uma lista com a ordem decrescente dos candidatos de acordo com a pontuação nas

perguntas e respostas. Se um eleitor deseja receber uma recomendação de voto para partidos, o procedimento é o mesmo.

Além disso, o *Smartvote* oferece ainda alguns serviços adicionais como uma base de dados com todos os perfis dos candidatos, como as respostas do questionário, sua carreira política, seus compromissos e background. O banco ainda inclui um *link* para o site pessoal de cada candidato.

O website do *Smartvote* ainda conta com algumas ferramentas visuais para facilitar a visualização do usuário com as preferências políticas dos candidatos de acordo com as respostas do candidato, utilizando gráfico de oito pontas (*smartspider*). O *smartspider* mostra a concordância ou discordância em oito grandes questões políticas formuladas como objetivos político.

2.1.3 Desafios da Democracia Eletrônica

Embora a democracia eletrônica tenha muitas vantagens como engajar as pessoas mais jovens a participar em maior quantidade, ajudar a melhorar a participação democrática, reduzir o analfabetismo cívico e a apatia dos eleitores, ainda existem alguns obstáculos limitadores como o fosso digital entre os que têm acesso aos meios de comunicação de e-democracia (telefones celulares e conexões de Internet), por exemplo, e os que não possuem, além dos investimentos tecnológicos que são necessários para a implantação bem sucedida do modelo [14].

Além disso, a exclusão digital e suas implicações para a igualdade política são zonas de perigo em potencial para a democracia eletrônica. O analfabetismo em TIC é um motivo para desencorajar a participação da população, já que os avanços na Internet podem potencialmente desanimar não-usuários a se adaptarem às novas tecnologias [14] [15]. É importante observar também que as principais vítimas da exclusão digital são as camadas menos favorecidas da sociedade, que ficariam desta forma excluídas de uma parte importante do processo democrático.

2.2 TECNOLOGIAS UTILIZADAS NA DEMOCRACIA ELETRÔNICA

Nesta seção serão apresentadas as principais tecnologias como Web 2.0, redes sociais e processamento de linguagem natural que compõem os sistemas utilizados para democracia eletrônica.

2.2.1 Web 2.0

No final do ano 2004, o termo Web 2.0 foi popularizado por Tim O'Reilly, fundador da O'Reilly, uma das principais organizações de mídia sobre tecnologia no mundo. Na conferência *O'Reilly Media Web 2.0*, que tinha como principal assunto a ser abordado "A Web como plataforma", Tim O 'Reilly citou este termo, mostrando como a Web poderia ser uma plataforma robusta para a inovação em muitos meios de comunicação e dispositivos.

Assim como em softwares, 2.0 significa uma melhoria, ou, uma nova geração da *World Wide Web* (WWW). Com isso, novas tecnologias e paradigmas fizeram com que o cidadão comum que utiliza a Internet passasse a colaborar e interagir. Na Web 2.0, o papel dos internautas muda drasticamente de um consumidor passivo de informação, para usuários que se tornam organizadores ativos de conteúdo online. De fato, cada vez mais indivíduos ou grupos optam por publicar suas músicas favoritas, textos ou vídeos on-line, compartilhando de suas mensagens em sites interativos [16].

A Web 2.0, conforme definida por Tim O'Reilly, "é uma tendência na utilização da tecnologia e design WWW na Web que visa facilitar a criatividade, a informação compartilhada e colaboração entre os usuários". Aplicações da Web 2.0 têm aparecido em grandes sistemas da Internet como o Google AdSense, Flikr, Napster, Wikipedia e Youtube. Também pode ser encontrado em Blogs, fórum, serviços Web, etiquetagem (*folksonomia*)[1] e distribuição de conteúdo [16] [17].

Alguns autores argumentam que a Web 2.0 pode ser vista como uma versão colaborativa da Web, ou seja, onde os usuários possam interagir. Por isso, no contexto de democracia eletrônica, é justamente essa característica da Web 2.0 que pode ser uma grande ferramenta para difundir e melhorar a participação do cidadão na tomada de decisão. A Web

[1] A **folksonomia** é uma maneira de indexar informações, através de palavras-chaves conhecidas como tags.

2.0 é uma coleção de tecnologias, negócios, estratégias e tendências sociais mais dinâmica e interativa do que o sua antecessora, a Web 1.0, permitindo aos usuários que acessem o conteúdo de um site e também contribuam com ele [17].

Com o crescente interesse em tecnologias Web 2.0 e aplicações, mais evidências mostram que a Web 2.0 ajuda as organizações públicas a oferecer um serviço de alto desempenho em termos de sua capacidade de gerar resultados sociais positivos, para servir ao bem comum, para envolver o público como coautores e melhorar a prestação de contas e transparência [18] [19].

Reconhecendo o enorme potencial na Web 2.0, muitos serviços públicos estão utilizando sites e redes sociais para estabelecer uma maior relação com os cidadãos além de conseguir um maior retorno. Deste modo, o governo consegue engajar uma maior gama de cidadãos como os jovens e também aqueles que vivem em áreas mais afastadas, atingindo um grande foco da democracia eletrônica, conhecido como *e-participation* [18].

Com todas funcionalidades que a Web 2.0 trouxe, tornando a plataforma Web muito mais participativa e colaborativa, algumas ferramentas baseadas nestes recursos e preceitos tem como finalidade auxiliar os gestores públicos e agências governamentais a melhor ouvir o cidadão. Desta maneira o cidadão pode passar a exercer um papel mais participativo, e realmente fazer parte das ações e planos que os governos fazem para a população, auxiliando na formação de políticas públicas e na tomada de decisão.

Utilizar alguns conceitos e tecnologias da Web 2.0 em questões de governança poderia implicar em uma mudança de políticas públicas, leis e serviços por parte das autoridades estabelecidas para seus clientes (povo). Petrik [19] ilustra como a Web 2.0 pode ser aplicada em um software descrito como um sistema de democracia eletrônica (EDS - E-Democracy System), envolvendo tanto o governo como o cidadão no processo de formulação de políticas. Este sistema iria buscar uma participação em conjunto para decisões consensuais informadas em uma forma deliberativa e colaborativa, livre de polêmica, meios de comunicação, preconceito e corrupção.

O sistema proposto pode ser descrito em seis fases que seguem abaixo:

1. Sistema de Sugestão: pessoas sugerem petições, problemas observados, metas políticas e os próprios usuários dão nota criando um *ranking*.

2. Lobby System: este serviço fornece ferramentas da Web 2.0 para que os cidadãos possam se comunicar, formar grupos com interesses comuns e organizar estratégias para tomada de decisões (redes sociais).
3. E-fórum: debater, votar, discutir propostas entre os cidadãos (usuários).
4. Wiki Política: contém informações da situação dos locais, fatos, detalhes e informações (Wiki sobre os assuntos).
5. Sistema de avaliação: avaliar como a proposta escolhida (1) vai afetar os cidadãos.
6. Sistema de tomada de decisão: votação na petição.

A seguir, é apresentado um guia para a aplicação do EDS no processo político que pode ser dividido em quatro fases: definição do problema, definição do objetivo da política, avaliação da opção, implementação e avaliação [19].

A primeira fase é a definição do problema, utilizando-se o sistema para engajar a população. Primeiramente, o cidadão relata um problema através do sistema de sugestão. Vários problemas podem ser relatados, mas são os próprios usuários (povo) que são atraídos pelo problema sugerido e que vão decidir quais são os problemas mais relevantes.

Com mais pessoas aderindo ao problema exposto, os usuários que observaram, ou estão tendo o mesmo problema, através do Lobby System podem se juntar e montar um grupo para trocarem informações de um modo mais fácil, utilizando ferramentas da web 2.0, como opções de *like*, além dos usuários contribuírem, ou seja, são eles próprios que colocam, validam, revisam e comentam.

Após formado um grupo de usuários na mesma causa, utilizando o e-fórum para debater ideias, os cidadãos conseguem juntar mais informações sobre o problema, como outros lugares também afetados e também sugerir soluções.

Utilizando a Wiki, os usuários podem gravar suas observações individuais, anexar fotos em um mapa digital, e particularizar a situação indesejável para a informação dos outros cidadãos, assim todos podem tomar conhecimento do problema e onde ele está ocorrendo, outra característica da Web 2.0, participação do usuário, os usuários do site adicionam conteúdo para os outros verem.

Após o problema ter sido identificado e um número mínimo exigido de cidadãos ter aderido ao debate do problema, os partidários decidem em uma pesquisa interna para propor um voto popular. Agora, o sistema de tomada de decisão é utilizado, assim os próprios cidadãos vão fazer valer seu voto, decidindo ou não, se o problema deve ser tratado como uma política pública.

Tendo definido um problema, a segunda fase é definir um objetivo para a política que vai ser adotada. Logo, todos os envolvidos deixam sua proposta para resolver o problema no sistema de sugestão e a proposta com maior aceitação é definida como o objetivo, levando à terceira fase: escolher a estratégia da política.

A terceira fase é a chave do ciclo político, aqui uma decisão sobre a ação política tem de ser determinada e o grau de participação direta atinge o seu pico. Neste ponto, o Sistema de Sugestão (1) é iniciado, agora, pelo governo que pede para que o público apresente estratégias para atingir o objetivo da política comum.

O Sistema de Avaliação (5) é aplicado para calcular o provável impacto de cada estratégia. Agora políticos analistas, empresas, os diferentes lobbies e outros especialistas determinam o lado positivo e negativo de cada estratégia usando outro aplicativo da Web 2.0. Como resultado, o sistema de avaliação mostra a força e as fraquezas de cada sugestão.

Na fase quatro, depois dos cidadãos definirem qual estratégia vai melhor resolver o problema, a estratégia é implementada, e a seguir, no final do ciclo de política, o eleitorado é solicitado para avaliar se o objetivo político foi alcançado, comparando à meta pré-definida.

Como podemos ver, o sistema EDS usa vários preceitos da Web 2.0. Entre eles, podemos citar a *folksonomia,* por meio da qual os usuários adicionam informação para os outros, e interagem por meio de avaliação, revisão e comentários. Além disso, as contribuições ficam disponíveis para todos.

Outro sistema de democracia eletrônica que utiliza técnicas e conceitos de Web 2.0 tem sido usado na Escócia. O The Highland Youth Voice é uma tentativa de aumentar a participação democrática, utilizando medidas on-line do Conselho de Highland para incentivar os jovens que vivem na região das montanhas da Escócia a participar da tomada de decisão democrática sobre suas próprias vidas. O Highland Youth Voice tem como

objetivo aumentar a participação da geração mais jovem através da compreensão de suas necessidades e desejos de seu governo, levando em conta seus pontos de vista, experiências e aspirações, dar voz aos jovens escoceses e chance de influenciar aqueles que tomam decisão [20].

Com auxílio do site, os envolvidos são capazes de discutir as questões importantes para eles. O aspecto democrático proeminente final do site são as eleições para os membros, que ocorrem a cada dois anos. Esses conteúdos do site permitem um fórum online em que os membros podem educar-se, participar em debates políticos on-line, ou experimentar um modelo de democracia eletrônica na facilidade de votação online [20].

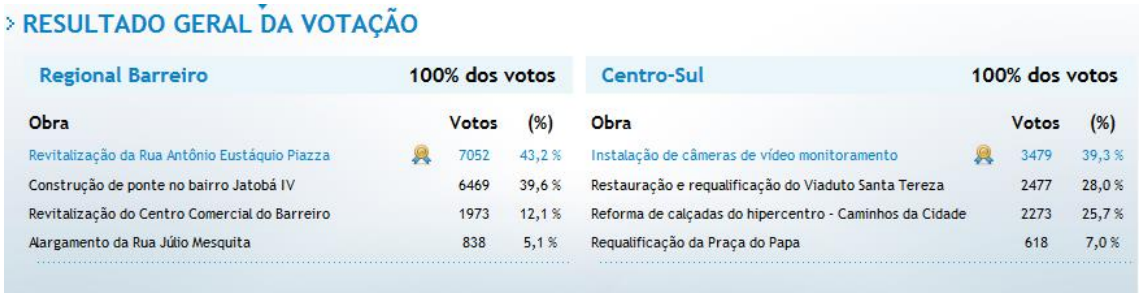
Continuando o projeto escocês, em 16 de maio de 2002, o Primeiro Ministro escocês Jack McConnell lançou um portal de informação para os jovens (12 - a 24 anos de idade) que vivem na Escócia (www.youngscot.org). Originalmente estabelecida há mais de 20 anos, a Young Scot fornece uma ampla gama de informações e serviços para jovens, incluindo manuais, boletins e o livre acesso aos conselhos. O canal Ur'Say inclui a criação de conteúdo do portal, desenvolvimento de novos modelos de facilitação on-line e moderação, e desenvolver a participação democrática e modelos de consultas aos jovens. Os fóruns de discussão do Ur'Say tem como objetivo facilitar o diálogo entre os jovens sobre questões sociais e políticas contemporâneas de relevância e importância para suas vidas [20].

Desta maneira, com o avanço da tecnologia, a interação, compartilhamento de informação, o modo como cada usuário pode contribuir e realmente iniciar os temas de discussão, bem como, opinar, validar outros temas, só foi possível com a utilização da web 2.0

Outras formas de abordagem também são usadas para engajar mais a população na tomada de decisão como em Belo Horizonte. O sistema de orçamento participativo e-PB (*e-Participatory Budgeting*) consiste em um esquema onde os cidadãos registrados como eleitores de Belo Horizonte, independente do seu local de residência na cidade, votam exclusivamente on-line para escolher uma de quatro obras públicas para cada um dos nove distritos da cidade para serem feitas.

Segundo a administração, o lançamento da iniciativa teve três objetivos principais: modernizar o seu orçamento participativo, através da utilização das TIC, aumentar a participação dos cidadãos no processo de orçamento participativo e alargar o âmbito das obras públicas que são submetidas à votação. O processo de e-PB começa com a seleção de

quatro obras públicas por distrito, que será sujeito à votação on-line com o objetivo de selecionar apenas um trabalho por distrito. Os trabalhos a serem selecionados para o e-PB visam responder às demandas de maior alcance que irão beneficiar uma parcela maior da população [21].



RESULTADO GERAL DA VOTAÇÃO

Regional Barreiro	100% dos votos		Centro-Sul	100% dos votos	
Obra	Votos	(%)	Obra	Votos	(%)
Revitalização da Rua Antônio Eustáquio Piazza	7052	43,2 %	Instalação de câmeras de vídeo monitoramento	3479	39,3 %
Construção de ponte no bairro Jatobá IV	6469	39,6 %	Restauração e requalificação do Viaduto Santa Tereza	2477	28,0 %
Revitalização do Centro Comercial do Barreiro	1973	12,1 %	Reforma de calçadas do hipercentro - Caminhos da Cidade	2273	25,7 %
Alargamento da Rua Júlio Mesquita	838	5,1 %	Requalificação da Praça do Papa	618	7,0 %

Figura 2.1 - resultado da obra vencedora em duas regiões [22]

A plataforma de e-voting é utilizada para que os cidadãos com mais de 16 anos de idade registrados em Belo Horizonte, possam votar na escolha de umas das obras, acessando a plataforma de e-voting através do site oficial da cidade. Para melhor descrever e compreender o funcionamento da plataforma de *e-voting*, quatro características principais são descritas a seguir: prestação de informações, interatividade bilateral, interatividade multilateral e ferramentas de e-voto [21].

A respeito do fornecimento de informações, o site fornece informações gerais sobre a iniciativa de orçamento participativo. Esta informação foi complementada com um link para "perguntas mais frequentes". Cada proposta foi indicada geograficamente e seguido por um pequeno texto padrão com uma descrição. Para cada proposta uma imagem foi associada, representando onde obras ocorreriam caso fosse selecionada. Esta descrição e imagem foram complementadas ainda por dois vídeos [21].

A Interatividade Bilateral é entendida como a característica da plataforma virtual e-PB que permite aos cidadãos terem acesso à informação (por exemplo, endereços de e-mail) e interagir com os membros da administração a cargo do e-PB, contatando-os diretamente [21].

Interatividade Multilateral é a característica que inclui não só a interação bilateral entre cidadãos e eleitos, mas também a possibilidade de debate oferecida por uma plataforma de votação, por exemplo, através de fóruns on-line, chats e listas de discussão [21].

A fim de implementar uma plataforma de votação (ferramenta de e-voto), a principal preocupação era garantir que o sistema fosse capaz de identificar fraudes ou tentativas de voto duplo, ou seja, só permitiria que eleitores com direito de tomar parte na votação um único voto [21].

Os resultados obtidos com a utilização da TIC para auxiliar o eleitor na escolha de como utilizar os recursos e onde utilizá-los, mostrou um aumento de sete vezes no número de participantes, e ainda utilizando um número bem inferior da renda se comparado com a maneira tradicional que era utilizada anteriormente como visto na Tabela 2.2 [21].

Tabela 2.2 - Participação do cidadão no Orçamento Participativo [14]

	Recursos	**Participação**
PB tradicional	US$ 43 Milhões	1.46%
e-PB	US$ 11 Milhões	9.98%

2.2.2 Redes Sociais

A democracia eletrônica e as mídias sociais são as bases de uma nova forma de engajamento entre o governo e a comunidade em geral. Muito mais do que uma forma de se relacionar online, a mídia social oferece ao governo a capacidade de se conectar e compartilhar ideias com o eleitorado [23].

O uso bem sucedido da Internet durante a campanha presidencial de 2008 nos EUA levantou novamente vozes que afirmam que aplicações como Twitter podem proporcionar novas oportunidades para a campanha online [24].

Levando em conta o grande sucesso da campanha eleitoral dos EUA de 2008, Larsson e Moe [16] fizeram um estudo sobre as eleições de 2010 suecas, utilizando o Twitter e uma ferramenta chamada TwapperKeeper para arquivar os tweets relevantes para a eleição e, posteriormente, realizar uma análise sobre os dados coletados [24].

A ferramenta TwapperKeeper permite que os usuários possam adquirir os *tweets*, que são as mensagens que cada usuário do Twitter publica, seguindo alguns critérios de filtros. Para realizar a análise posterior, o texto, nome de usuário e ID de mensagem do remetente, ID de usuário do destinatário (se a mensagem é uma resposta), o código de linguagem, o cliente usado para enviar *tweet*, código geográfico, e o tempo que o *tweet* foi

criado, foram armazenados. Com estes dados organizados em uma base de dados, algumas análises foram feitas com a utilização de uma ferramenta grátis para estatística (SPSS) e uma ferramenta grátis para grafos (Gephi) [24].

Para ter uma boa base de informações sobre as eleições, foram capturados tweets relacionados à eleição de um mês anterior a realização da mesma e três dias posterior ao término do pleito eleitoral, tendo seu pico no dia em que foi realizada a eleição [24].

De acordo com os autores deste estudo, o Twitter contribuiu claramente para a ampliação do debate como um novo meio para a comunicação pública. No entanto, a relação dos tweets com o que está acontecendo no momento sinaliza a relação entre mídia e Twitter: picos de atividade em tweets sobre a eleição podem ser ligados a quaisquer eventos como talk-shows televisivos, ou a cobertura da mídia de eventos off-line, como comícios políticos [24].

Na seção 3, faremos um experimento utilizando a mineração de dados para extrair dados relevantes das informações coletadas do Twitter.

2.2.3 Processamento de Linguagem Natural (PLN)

Processamento de Linguagem Natural (PLN) é um ramo da inteligência artificial que trata de analisar, compreender e gerar as linguagens que os seres humanos usam naturalmente, tanto escrita como falada, a fim de interagir com computadores, utilizando as línguas humanas naturais em vez de linguagens de computador.

A história da PLN teve seus primórdios no começo dos anos 50, apesar de existirem alguns relatos anteriores do final da década de 40. Em 1950, Alan Turing, publicou um artigo intitulado "Computing Machinery and Intelligence", que propôs o que agora é chamado de teste de Turing como um critério de inteligência [25].

Devido à grande tensão existente entre Estados Unidos e União Soviética, um experimento conhecido como Experimento Georgetown, em 1954, envolveu a tradução automática de mais de sessenta frases do russo para o inglês usando o IBM-701. A partir dos anos 90, o PLN começou a crescer mais rápido do que nunca [25].

Em particular, as grandes quantidades de texto inundando a World Wide Web (WWW) estimularam o trabalho em mecanismos para a gestão deste fenômeno, nomeadamente através da extração de informações e resumo automático. A criação e uso público da Internet juntamente com enormes quantidades de textos em Francês e Inglês ajudaram na revitalização de aprendizado de máquina e, portanto, a tradução automática. Com toda esta nova informação e textos legíveis, houve um grande avanço no uso da linguagem falada e reconhecimento de fala [25].

O PLN tem como objetivo desenvolver modelos de como o ser humano compreende e produz linguagem natural para construir som e fazer teorias sobre a relação entre a mente e a linguagem a fim de melhorar e facilitar a interação humano-computador, que é a PLN aplicada, e mostrar que o processamento de linguagem natural pode ser um instrumento poderoso para acesso à informação e assim melhorar a interação entre os usuários através de computadores.

Para melhorar a interação entre os usuários através de computadores, o projeto Rede Europeia de Democracia Eletrônica (EDEN), visa descobrir se uma abordagem da PLN em particular pode promover a democracia eletrônica aumentando a participação dos cidadãos no processo de tomada de decisões. [26]

A equipe de desenvolvimento do projeto EDEN desenvolveu um software dividido em cinco módulos que foram baseados em tecnologias PLN e estruturados de acordo com os usuários (administradores públicos) e as áreas em que eles trabalham. [26]

- Address Guesser: automaticamente seleciona o administrador público responsável de acordo com o conteúdo da mensagem.
- Answer Tree: responsável pelas FAQ (perguntas frequentemente feitas) e recupera as perguntas e respostas mais semelhantes à pergunta do usuário.
- Style Enhancer: que analisa um documento para melhorar sua legibilidade antes do administrador público publicá-lo em seu site.
- Multilanguage Helper: liga termos técnicos às suas traduções em diferentes linguagens, proporcionando definições simples.

- Natural-Language Map: recupera textos e mapas através de uma interface de linguagem natural.

Um sistema de avaliação sobre o EDEN foi feito em algumas cidades europeias e seus resultados de como a ferramenta se saiu, bem como cada parte da ferramenta pode ser encontrado em [27].

3 DEMOCRACIA ELETRÔNICA E MINERAÇÃO DE TEXTO NO TWITTER

Neste capítulo será mostrado como funciona o processo de mineração de dados, tendo como fonte dos dados que serão analisados o Twitter, e assim mostrar como este tipo de abordagem pode contribuir para o processo de democracia eletrônica.

O Twitter frequentemente reflete os pensamentos e vontades momentâneas de uma boa parte da população que tem acesso à Internet. Desta forma, o Twitter pode ser também uma interessante ferramenta para os governantes, considerando que qualquer político ou administração que não se alimenta das ideias de seus administrados está deixando passar uma grande oportunidade de saber o que a população pensa sobre seu mandato [28]. As técnicas de mineração de texto podem, por sua vez, ajudar os analistas a obter informações importantes do grande volume de dados que é disponibilizado nas redes sociais.

3.1 TWITTER

Redes sociais como o Orkut, Facebook e LinkedIn são aplicações Web que permitem a criação e troca de dados gerados pelos usuários [29]. O Twitter, por sua vez, é um microblog, que é uma forma de publicação que permite aos usuários fazer atualizações breves de texto incluindo, em geral, reflexões curtas ou informações momentâneas do dia-dia. Assim sendo, plataformas de microblogging [30] como o Twitter são uma ótima maneira de ver como as pessoas reagem a diferentes assuntos e temas [29] [31].

O Twitter foi criado em 2006 e tornou-se o serviço de microblog mais popular atualmente, onde é possível ao usuário enviar mensagens de no máximo 140 caracteres, conhecidas como *tweets*. Os usuários utilizam esses *tweets* não só como um meio para divulgar suas ideias e sentimentos, mas também para expressar sua opinião sobre produtos, serviços e eventos [31].

No Twitter, o usuário possui seguidores e também pode seguir outras pessoas, criando sua rede de amizades e interesses. Há ainda a possibilidade de enviar mensagens em modo privado para outros perfis, além de divulgar vídeos, fotos e links.

Em uma pesquisa recente realizada no meio do ano de 2013, o Twitter revelou que teve 218,3 milhões de usuários ativos mensais médios durante três meses. A empresa afirmou ainda que tem 215 milhões de usuários ativos mensais, além de contar com 100 milhões de usuários ativos por dia, e um número de 500 milhões de *tweets* diários [32].

No mesmo período de tempo, o Twitter descobriu que 75% desses acessos são a partir de um dispositivo móvel, o que equivale a 163.500.000 usuários móveis ativos mensais [32]. É importante salientar que o Twitter é uma rede social "location-aware", pois permite aos usuários móveis mostrar sua geo-localização. A capacidade de coletar e extrair os tweets que possuem geo-localização abre novas direções para pesquisas [33].

Um aspecto importante do Twitter é que ele permitiu que qualquer pessoa com acesso a um celular que possa se conectar à Internet se comunique rapidamente, possibilitando a um cidadão comum que testemunhe algum evento compartilhar e espalhar a notícia, assim podendo alertar e ajudar durante emergências de incêndios, a situação de como o tráfego está em determinado trecho, como engarrafamentos, acidentes, além de outros eventos [33].

Desta forma, as grandes quantidades de dados públicos que fluem através das redes sociais tem o potencial de oferecer uma nova visão valiosa para o comunidade acadêmica, agências de marketing, ONGs e outras organizações interessadas em entender o comportamento on-line e monitoramento [33].

3.2 MINERAÇÃO DE TEXTO

Neste trabalho, o objetivo é investigar como podemos aproveitar o volume de informações que são submetidas no Twitter para extrair informações úteis para democracia eletrônica, aproximando os governantes da população. Para extrair informações do Twitter que possam ser de alguma relevância para gestores ou órgãos públicos, podemos usar técnicas de mineração de texto, que se dedicam a derivar informações relevantes de um grande conjunto de textos [34].

O processo de mineração de texto é obtido através da elaboração de padrões e tendências baseados em técnicas de aprendizagem de padrões estatísticos. Mineração de texto e mineração de dados são parte de um processo maior conhecido como KDD

(Knowledge Discovery in Databases) – em português, Descoberta de Conhecimento em Bases de Dados –, que, segundo [35], permite a extração não trivial de conhecimento previamente desconhecido e potencialmente útil de um banco de dados.

As informações obtidas através do processo de descoberta de conhecimento podem ser encontradas em três formas. Primeiro a informação estruturada, como em um banco de dados, segunda forma semiestruturada, quando armazenada em um documento XML, ou finalmente não ter nenhuma estrutura, e um exemplo desta forma é quando a informação está em um simples texto [36].

A grande diferença entre mineração de texto e mineração de dados é a diferença quanto a forma que as informações são encontradas. Na mineração de dados, as informações são encontradas de um modo mais estruturado como um banco de dados, já na mineração de texto é aplicada sobre informações armazenadas sem nenhuma estruturação como textos e documentos [37].

Mineração de textos é a procura por padrões em um texto em linguagem natural e pode ser definido como o processo de análise do texto para extrair informação dele para um propósito em particular.

O processo de mineração de texto envolve um mecanismo de coleta de dados, uma etapa de pré-processamento, que envolve indexação, aplicação do algoritmo (mineração de dados) e finalmente a análise dos resultados como podemos analisar na Figura 3.2 [38] [39].

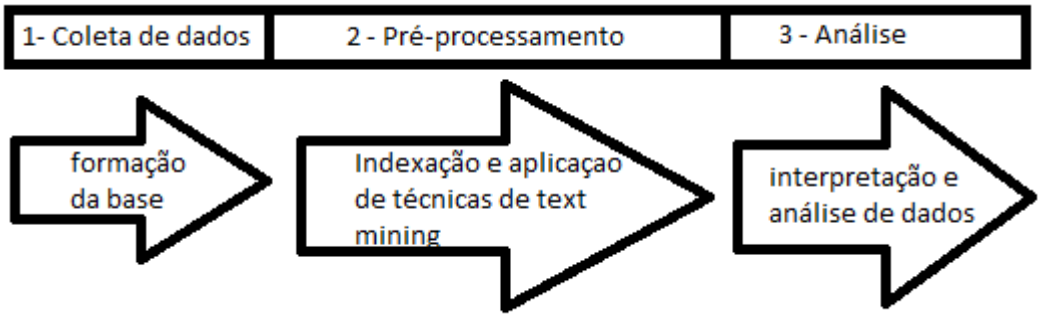


Figura 3.1 - Etapas da mineração de texto

3.2.1 Etapas da Mineração de Texto

A primeira etapa da mineração de texto, é chamada de coleta de dados, e corresponde a formação da base de textos que será usada para se extrair as informações. Este processo de coletar a informação se dá em duas formas, podendo ser uma base estática que, coleta as informações e armazena num banco, ou dinâmica que recebe atualização em tempo real, recebendo novos dados a todo momento [38] [39].

O processo da coleta de dados dá-se em três formas: sistemas de arquivos, banco de dados ou pela Web. Apesar dos três tipos diferentes, um sistema crawler é responsável por essa extração dos dados [39].

Após a formação da base de dados que será utilizada para a mineração de texto, vem a segunda etapa, que é chamada de pré-processamento e tem como objetivo melhorar a qualidade dos dados disponíveis e organizá-los de uma forma mais fácil para a análise.

As medidas adotadas nesta fase tem como finalidade submeter os dados a um processo de indexação e/ou aplicação de técnicas de mineração de texto. O pré-processamento significa dividir o texto em palavras, aplicação de técnicas de stemming (normalização morfológica), remover as stop-words e classificá-las segundo a classe gramatical [38] [39].

A remoção das stop-words é uma das etapas mais utilizadas no pré-processamento, pois são palavras comuns, geralmente artigos, pronomes, preposições, advérbios e outras palavras que não possuem valor semântico nenhum ao texto, ou seja, não agregam valor, assim reduzindo o tamanho dos textos, melhorando a indexação [38] [39].

O stemming por sua vez visa a remoção dos sufixos e prefixos, tomando apenas os radicais das palavras. Além disso, o stemming também é utilizado para remoção de características de gênero, número e grau das palavras, o objetivo desta técnica é a obtenção de um elemento único que possa ser considerado como um termo só [38] [39].

Existem algumas técnicas de stemming bastante utilizadas, como método do stemmer S, que basicamente elimina plurais, método de Porter que elimina vários tipos de sufixos e o método de Lovins, que elimina cerca de 250 sufixos [39].

A primeira etapa da mineração de texto, é chamada de coleta de dados, e corresponde a formação da base de textos que será usada para se extrair as informações. Este processo de coletar a informação se dá em duas formas, podendo ser uma base estática que,

coleta as informações e armazena num banco, ou dinâmica que recebe atualização em tempo real, recebendo novos dados a todo momento [38] [39].

O processo da coleta de dados dá-se em três formas: sistemas de arquivos, banco de dados ou pela Web. Apesar dos três tipos diferentes, um sistema crawler é responsável por essa extração dos dados [38].

Após a formação da base de dados que será utilizada para a mineração de texto, vem a segunda etapa, que é chamada de pré-processamento e tem como objetivo melhorar a qualidade dos dados disponíveis e organizá-los de uma forma mais fácil para a análise.

As medidas adotadas nesta fase tem como finalidade submeter os dados a um processo de indexação e/ou aplicação de técnicas de mineração de texto. O pré-processamento significa dividir o texto em palavras, aplicação de técnicas de stemming (normalização morfológica), remover as stop-words e classificá-las segundo a classe gramatical [38] [39].

A remoção das stop-words é uma das etapas mais utilizadas no pré-processamento, pois são palavras comuns, geralmente artigos, pronomes, preposições, advérbios e outras palavras que não possuem valor semântico nenhum ao texto, ou seja, não agregam valor, assim reduzindo o tamanho dos textos, melhorando a indexação [38] [39].

O stemming por sua vez visa a remoção dos sufixos e prefixos, tomando apenas os radicais das palavras. Além disso, o stemming também é utilizado para remoção de características de gênero, número e grau das palavras, o objetivo desta técnica é a obtenção de um elemento único que possa ser considerado como um termo só [38] [39].

Existem algumas técnicas de stemming bastante utilizadas, como método do stemmer S, que basicamente elimina plurais, método de Porter que elimina vários tipos de sufixos e o método de Lovins, que elimina cerca de 250 sufixos [39].

3.2.2 TF-IDF

O TF-IDF é uma estatística numérica que reflete o quão importante uma palavra é para um documento, coleção ou corpus [40]. Ele é frequentemente usado como um fator de peso na recuperação de informação e mineração de texto.

O TF (Frequência do Termo) é o número de vezes que uma palavra ocorre em um determinado documento. O IDF (Frequência Inversa Documento) de um palavra é baixo se ele ocorre em muitos documentos. Por outro lado, se a palavra ocorre em apenas um documento, seu IDF é alto [41].

TF-IDF é o produto de duas estatísticas, o TF e o IDF. Para se calcular o TF-IDF então é necessário encontrar primeiro os valores dos dois. No caso do TF, deve-se calcular a frequência total de um termo em um documento, ou seja, o número de vezes que o termo ocorre em um determinado documento e dividir pelo total de palavras que este documento possui.

O IDF é uma medida para saber se o termo é comum ou raro em todos os documentos. Para se obter o IDF, basta dividir o número total de documentos pelo número de documentos que contêm o termo, e, em seguida, tomar o logaritmo deste quociente como pode ser visto em (1), onde t é o termo (palavra), d é o documento e D o número total de documentos no corpus.

$$\boldsymbol{idf(t, D) = \log \frac{|D|}{|\{d \in D : t \in d\}|}} \qquad (1)$$

Conforme mencionado anteriormente, para se obter o resultado do TF-IDF é preciso ter antes os valores do TF e do IDF. Assim sendo, com os resultados já obtidos basta apenas dividir o valor do TF pelo IDF, que assim se obtém o valor final do TF-IDF.

O TF-IDF pode ser usado no Twitter para ranquear as palavras mais significativas em um conjunto de *tweets*, por exemplo. Como os *tweets* podem reunir discussões que sejam do interesse de governantes, ao utilizar o TF-IDF como ponderador de valor das palavras que aparecem, podemos ranquear as palavras mais significativas e assim extrair de todo o conjunto de *tweets* dados relevantes que podem ser usados por gestores públicos.

3.3 LEVANTAMENTO DE PALAVRAS MAIS SIGNIFICATIVAS EM *TWEETS* DIRECIONADOS A GOVERNANTES

Nesta seção, será apresentado um estudo realizado a partir da aplicação do TF IDF sobre tweets que tinham importantes governantes do Brasil como destino. O objetivo é elencar as 10 palavras mais significativas de *tweets* endereçados aos prefeitos de São Paulo, Rio de Janeiro e Curitiba e aos governadores do Paraná e do Rio de Janeiro e analisar a relação delas com fatos ocorridos durante o período de coleta.

Para acessar os tweets, foi utilizada a Twitter API. Os desenvolvedores do Twitter disponibilizam uma API (*Application Programming Interface*) REST (*Representational State Transfer*) que permite aos demais desenvolvedores acessar algumas das principais primitivas da rede social, incluindo tweets, atualizações de status e informações do usuário. Além disso, é possível construir um perfil de um usuário, utilizando seu nome, seu avatar de perfil e os demais usuários que ele segue [42].

A API (*Application Programming Interface*) do Twitter permite que os desenvolvedores criem aplicativos usando dados básicos do Twitter. A plataforma API fornece todas as funções disponíveis no próprio site do Twitter (twitter.com), tais como a Search API para consultar e filtrar o conteúdo de mensagens e Stream API que permite um monitoramento em tempo real dos tweets, sendo muito utilizada para mineração de texto [23].

O Twitter baseia sua API em REST (*Representational State Transfer*). A arquitetura REST refere-se a um conjunto de princípios de design de rede que definem recursos e formas de abordar e acessar dados, o que significa que o serviço funciona com a maioria dos formatos de distribuição na Web [42].

Arquiteturas REST convencionalmente consistem em clientes e servidores. O cliente inicia as solicitações para o servidor, logo o servidor faz o processo da solicitação requisitada e assim retorna a resposta adequada ao cliente, ou seja através da API se faz uma requisição HTTP para o Twitter, que processa a solicitação e retorna para o cliente.

Como pode-se ver na Figura 3.2, o cenário de funcionamento da coleta de dados e formação do banco de dados é realizada da seguinte forma. Primeiramente a aplicação Java se conecta com o Twitter para extrair os dados através da API de *stream*, que faz um monitoramento de cerca de 1% [43] do fluxo em tempo real dos *tweets* enviados. Para adequar ao projeto em questão, um filtro foi criado na aplicação para apenas receber os *tweets* dos governantes e prefeitos de algumas cidades e estados do Brasil.

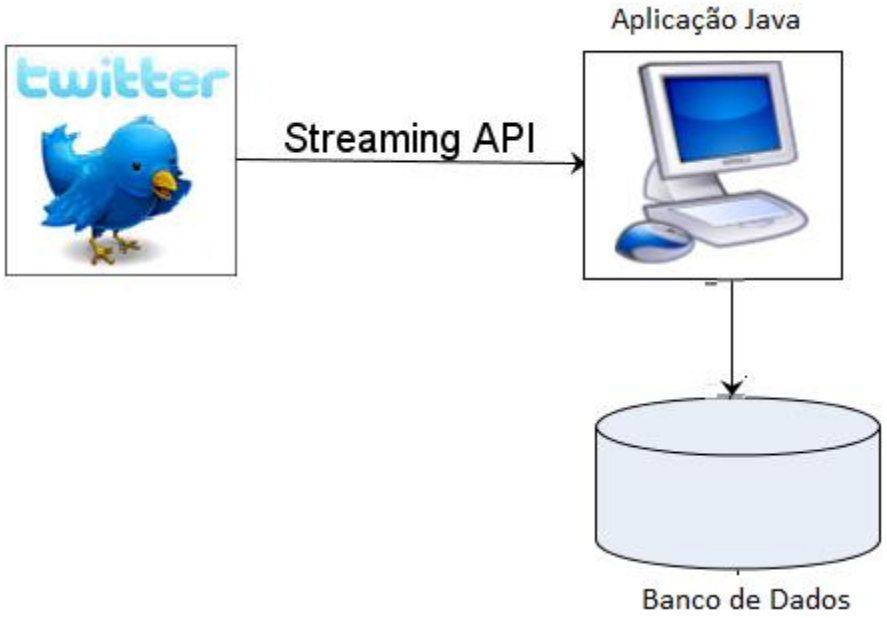


Figura 3.2 - Cenário da Aplicação

Os *tweets* que correspondem ao filtro são enviados a um banco de dados que está conectado a aplicação. O banco de dados foi construído em um SGBD PostgreSQL. O modelo de dados do banco é apresentado na Figura 3.3.

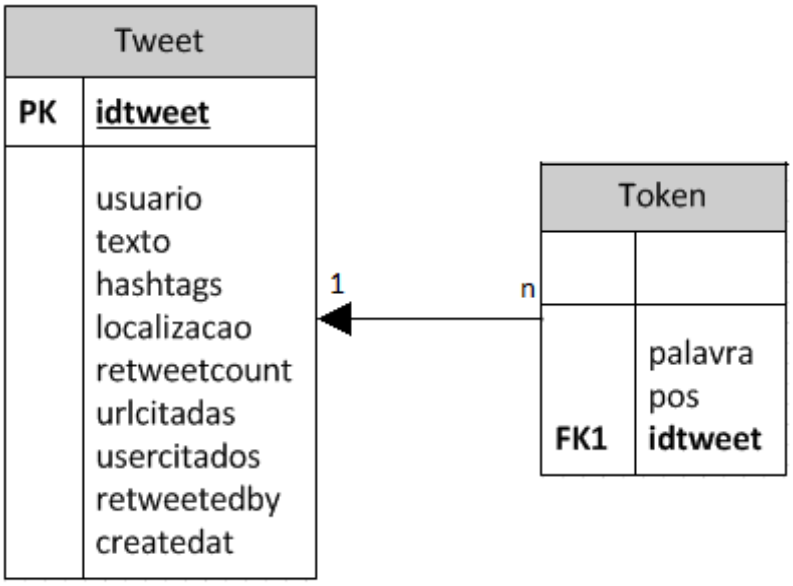


Figura 3.3 - Modelo de entidade relacionamento do banco usado

O modelo da Figura 3.3 mostra os campos das duas tabelas que foram usadas. A tabela *Tweet* contém todas as informações que um *tweet* possa ter, como quem

enviou, o que enviou, de onde o usuário enviou, além de outras características. Já a tabela Token armazena separadamente cada palavra presente em cada tweet.

A coleta dos tweets que mencionavam as contas no Twitter dos governantes foi realizada no período de 24 de Outubro a 31 de Outubro. Os governantes escolhidos foram Fernando Haddad, prefeito de São Paulo, Eduardo Paes, prefeito do Rio de Janeiro, Gustavo Fruet, prefeito de Curitiba, Sérgio Cabral, governador do Rio de Janeiro e Beto Richa, governador do Paraná. Ao final da coleta, a base de dados reunia aproximadamente 6500 *tweets*.

No pré-processamento da coleção de dados, utilizando-se uma base de stop-words, que pode ser vista na seção Anexo A previamente definida, alguns termos de irrelevância semântica foram retirados, como preposições, artigos entre outros.

Tendo agora os *tweets* armazenados, a próxima parte a ser feita foi a análise para levantar os termos mais relevantes. Para tanto, foi utilizado o TF-IDF.

O valor TF-IDF aumenta proporcionalmente ao número de vezes que uma palavra aparece no documento, mas é compensado pela frequência da palavra no corpus, que ajuda a controlar o fato de que algumas palavras são geralmente mais comuns do que outras. Portanto, o primeiro passo foi organizar os textos (*tweet*), em palavras únicas (*token*) e colocá-las no banco em outra tabela, como pode ser visto na Figura 3.3 utilizando um recurso da própria linguagem Java, chamado *Tokenizer*.

O cálculo do IDF foi feito sem nenhuma dificuldade, mas, para o valor do TF a abordagem inicial se mostrou problemática, pois como os *tweets* são limitados no máximo 140 caracteres a maioria dos valores encontrados é o mesmo, pois em um único texto não há grande variância de palavras utilizadas.

Para resolver o problema do TF, foi utilizada uma nova abordagem. Ao invés de utilizar os *tweets* separados, passou-se a usar eles em grupos. Cada grupo reúne todos os tweets de um determinado governante.

Assim foram ranqueadas as 10 primeiras palavras mais significativas de cada governante como segue abaixo nas tabelas. As Tabelas 3.1, 3.2, 3.3, 3.4 e 3.5 trazem as palavras mais relevantes dos *tweets* que mencionam Fernando Haddad (prefeito de São Paulo), Eduardo Paes (prefeito do Rio de Janeiro), Sérgio Cabral (governador do Rio de

Janeiro), Gustavo Fruet (prefeito de Curitiba) e Beto Richa (governador do Paraná), respectivamente.

Como é possível ver na Tabela 3.1, algumas palavras refletem fatos importantes ocorridos no período de coleta. Na semana em que foi realizada a coleta, o prefeito de São Paulo anunciou o aumento de IPTU (Imposto Territorial Predial Urbano) na cidade. Palavras como "Aumento", "Aumentar" e "IPTU" constam no ranking das 10 palavras mais significativas e, provavelmente, têm relação com este fato [44]. Entretanto, não é possível, a partir destas 10 palavras, definir qual foi a reação do público a esse aumento, já que encontramos entre elas a palavra "Obrigado" e a palavra "#foraPT" (uma menção ao partido do prefeito). É importante salientar que trabalhos futuros podem fazer análises mais profundas deste conjunto de palavras para tirar conclusões mais precisas sobre a opinião do público que usou o Twitter. Com base nos resultados obtidos com o TF-IDF, podemos apenas afirmar que três das 10 palavras tem forte relação com um importante fato do período, demonstrando que o Twitter refletiu de alguma forma decisões da administração pública.

Tabela 3.1 - Prefeito de São Paulo

Palavra	TF	IDF	TFIDF
IPTU	0,0175	2,9532	0,0059
Obrigado	0,0095	3,2631	0,0029
Paulo	0,0061	2,0128	0,0044
Aumento	0,0058	3,2812	0,0018
Prefeito	0,0056	3,1682	0,0017
#foraPt	0,0046	3,4767	0,0013
Periferia	0,0043	3,5733	0,0012
Parabéns	0,0043	3,2457	0,0012
Higienópolis	0,0041	3,7230	0,0011
Aumentar	0,0036	3,7230	0,0010

Analisando a Tabela 3.2, do prefeito do Rio de Janeiro Eduardo Paes, podemos observar um grande número de tweets de dois assuntos que movimentaram a semana de coleta. O primeiro deles foi o PCCS (Plano de Carreira, Cargos e Salários), motivo de grande divergência entre os professores e a prefeitura que gerou vários protestos [45]. O outro assunto foi a inauguração da Via Binária, prevista inicialmente para o fim de setembro e realizada no final de outubro [46].

Tabela 3.2 - Prefeito do Rio de Janeiro

Palavra	TF	IDF	TFIDF
Dúvidas	0,0050	3,1290	0,0016
Substituição	0,0046	3,1643	0,0014
Servidores	0,0044	3,6342	0,0012
Perimetral	0,0044	2,8175	0,0015
Binário	0,0044	3,1717	0,0013
Plano	0,0044	3,3119	0,0013
Encaminhamento	0,0044	3,6131	0,0012
Mande	0,0044	3,2362	0,0013
Abertura	0,0044	3,1290	0,0014
#PCCS	0,0044	3,6342	0,0012

Na tabela 3.3, temos as palavras mais significativas coletadas nos tweets que mencionavam a conta no Twitter do governador do Rio de Janeiro Sérgio Cabral. As palavras “IPVA” e “Devolução” refletem a decisão do governo do estado de devolver o IPVA ao contribuinte caso ele tenha seu veículo roubado ou ocorra perda total [47]. As palavras “Super-poderes”, “Cedae”, “Revogue” e “Presidente” provavelmente estão relacionadas a um decreto do governador que aumentou os poderes do presidente da CEDAE (Companhia Estadual de Águas e Esgotos) [48].

Tabela 3.3 - Governador do Rio de Janeiro

Palavra	TF	IDF	TFIDF
Decreto	0,0172	3,1027	0,0055
Revogue	0,0172	3,1027	0,0055
Cedae	0,0165	3,1156	0,0053
Super-poderes	0,0162	3,1290	0,0052
Presidente	0,0147	2,9758	0,0049
#SaneamentoÉBásico	0,0119	3,2631	0,0036
Assine	0,0119	3,2631	0,0036
Caso	0,0071	3,3669	0,0021
IPVA	0,0055	3,5550	0,0015
Devolução	0,0053	3,6131	0,0014

Na semana de coleta de dados no Twitter, um importante fato ocorrido na cidade de Curitiba foi a liberação de recursos para a construção da linha de metrô [49]. Na Tabela 3.4, que apresenta as palavras mais significativas encontradas nos tweets que

mencionavam o prefeito de Curitiba, podemos ver duas palavras relacionadas a este assunto: “Metrô” e “Recursos”.

Tabela 3.4 - Prefeito de Curitiba

Palavra	TF	IDF	TFIDF
Curitiba	0,0106	2,5658	0,0041
Prefeito	0,0089	3,0161	0,0029
Maior	0,0065	3,03742	0,0021
Mais	0,0058	3,5550	0,0023
Cidade	0,0056	2,9618	0,0019
Momento	0,0054	3,0542	0,0017
Inter	0,0052	2,9014	0,0018
Metrô	0,0052	2,9395	0,0017
Recursos	0,0047	3,2819	0,0014
Revitalização	0,0041	3,6131	0,0011

Para o governador do Paraná, cujos resultados são apresentados na tabela 3.5, o nome da presidente Dilma foi o termo que mais ocorreu e este fato pode estar relacionado a verba anunciada pela presidente para a construção da linha de metrô em Curitiba [50].

Tabela 3.5 - Governador do Paraná

Palavra	TF	IDF	TFIDF
Dilma	0,0350	3,0374	0,0115
Mente	0,3216	3,8213	0,0084
Anuncia	0,3216	3,4438	0,0093
Obra	0,0292	3,6800	0,0079
Paraná	0,0292	2,7592	0,0077
Denuncia	0,0292	3,8561	0,0075
Verba	0,0292	2,5928	0,0081
Pres	0,0116	4.2363	0,0027
Tipo	0,0116	4.2363	0,0027
IPI	0,0116	4.2363	0,0027

Ao observar as palavras apresentadas nas tabelas e a relação delas com os fatos importantes ocorridos no período, pode-se observar que os *tweets* que mencionavam os governantes refletiram os assuntos que mais mobilizaram uma parcela da população. Logicamente, não podemos afirmar que o Twitter reflete com exatidão os assuntos mais importantes a serem tratados pelos governantes, mas os resultados mostram que o

aprofundamento na aplicação de mineração de texto pode trazer importantes indicadores sobre o impacto de determinadas decisões políticas sobre a população. Em suma, como as redes sociais acabam refletindo os anseios momentâneos da parcela da população que tem acesso à Internet, acreditamos que elas podem ser um importante veículo para a democracia eletrônica.

4 CONCLUSÃO

A área da democracia eletrônica é ainda recente e cada vez mais estudos estão sendo feitos para encontrar soluções e novas técnicas para que as TIC possam colaborar para que os cidadãos participem cada vez mais do processo político, na formulação de novas leis e ideias para o governo, tornando o processo político mais transparente.

Entre as técnicas aplicadas à democracia eletrônica temos a Web 2.0 e o processamento de linguagem natural. A Web 2.0 permite que os usuários da web possam colaborar com a produção de conteúdo. Isso traz maior interatividade, característica muito importante para a democracia eletrônica. O processamento de linguagem natural pode ser utilizado para tornar a comunicação entre o gestor público e o cidadão melhor e mais ágil.

Além disso, as redes sociais, como o Twitter, podem ser uma boa alternativa para se extrair informações dos cidadãos, já que as pessoas expressam suas opiniões e relatam fatos do cotidiano nas redes sociais. O experimento realizado neste trabalho e apresentado na seção 3.3 demonstra isso. Neste experimento, o algoritmo de mineração de texto TF IDF foi aplicado para identificar automaticamente as 10 palavras mais significativas dos tweets endereçados a importantes governantes como o prefeito de São Paulo e do Rio de Janeiro. Nos resultados, foi possível observar que algumas das palavras consideradas mais significativas tiveram relação direta com importantes fatos observados no período.

Os trabalhos futuros abordam o aprofundamento da utilização de ferramentas de mineração de texto para identificar de maneira mais completa e automática quais são os temas mais abordados pelos usuários do Twitter e qual a opinião deles sobre estes temas.

5 REFERÊNCIAS BIBLIOGRÁFICAS

[1] Guo, Y. (2011) “Analysis on How to Enhance E-Democracy through E-Government”, In: Management and Service Science (MASS), 2011 International Conference on 12-14 Aug. 2011.

[2] Guimarães T. A., Medeiros P. H. R., “A relação entre governo eletrônico e governança eletrônica no governo federal brasileiro”, www.ebape.fgv.br/cadernosebape Volume III, Número 4, Dezembro 2005.

[3] Lee, C., Chang, K. and Berry, F.S (2011) “Testing the Development and Diffusion of E-Government and E-Democracy: A Global Perspective”, In: Public Administration Review vol 71, Issue 3, p 444–454, May/June 2011.

[4] Sigala, M. and Marinidis, D. (2010) “DMOs, e-Democracy and Collaborative Destination Management: an Implementation Framework” In: ENTER, p 235-246. Springer Vienna, (2010).

[5] Coleman S. and Norris D. (2005) “A new agenda for e-democracy”, In: Oxford Internet Institute, Forum Discussion Paper No. 4, January 2005.

[6] Emmanouil, S. and Konstantina, T. (2011) “Evaluating the Quality of e-Democracy Processes: an Empirical Study in the Greek Context” In: University of Macedonia Thessaloniki, May 2011.

[7] Ludlow, D. and Khan, Z. (2012) “Participatory Democracy and the Governance of Smart Cities”, In: University of the West of England, Bristol, UK, july 2012.

[8] Chen, H. (2009) “AI, E-government, and Politics 2.0”, In: Intelligent Systems, IEEE, p. 64-85, setember 2009.

[9] Päivärinta, T. and Sæbø, O. (2006) “Models of e-Democracy” In: Communications of the Association for Information Systems, vol 17, number 1, article 37, june 27.

[10] Abreu, J.C.A; Melo, D.R.A; Motta, G.S. “Modelos de Democracia Eletrônica: Analisando o Orçamento Participativo Digital”. Encontro de Administração Pública e Governança. Vitória/ES – 28 a 30 de novembro.

[11] Disponível em <http://bd.camara.gov.br/bd/bitstream/handle/bdcamara/12096/democracia_eletronica_stabile.pdf?sequence=1>. Acessado em 27 de Novembro de 2013 às 19:55.

[12] How direct democracy makes Switzerland a better place. Disponível em <http://www.telegraph.co.uk/news/1435383/How-direct-democracy-make s-Switzerland-a-better-place.html>. Acessado em 27 de Novembro de 2013 às 20:12.

[13] Fivaz, J. and Schwarz, D. (2007) "Nailing the Pudding to the Wall – E-Democracy as Catalyst for Transparency and Accountability" In: International Conference on Direct Democracy in Latin America; 14-15 Mar. 2007; Buenos Aires, Argentina.

[14] Benefits and Disadvantages. Disponível em <http://edemocracy1.blogspot.com.br/2011/02/benefits-and-disadvantages.html>. Acessado em 28 de Novembro de 2013 às 14:13.

[15] Promise and Problems of E-Democracy Disponível em <http://www.oecd.org/governance/public-innovation/35176328.pdf>. Acessado em 28 de Novembro de 2013 às 14:49.

[16] Breindl, Y. and Francq, P. (2008) "Can Web 2.0 applications save e-democracy? A study of how new internet applications may enhance citizen participation in the political process online", Int. J. Electronic Democracy, Vol. 1, No. 1, 2008.

[17] Murugesan, S. (2007) "Understanding Web 2.0" In: IT Professional vol: 9, Issue: 4, IEEE, p. 34-41, July-Aug. 2007.

[18] Tsui, H. D., Lee, C. Y. and Yao, C.B. (2010) "E-government, politics and Web 2.0", In: New Trends in Information Science and Service Science (NISS), 4th International Conference, p. 687-691, 11-13 May 2010.

[19] Petrik, K. (2009) "Participation and e-Democracy How to utilize Web 2.0 for policy decision-making", In: The Proceedings of the 10th International Digital Government Research Conference.

[20] Macintosh, A., Robson, E., Smith, E. and Whyte, A. (2003) "Electronic Democracy and Young People", In: Social Science Computer Review February 2003 vol. 21 no. 1, p 43-54.

[21] Peixoto, T. (2008) "e-Participatory Budgeting: e-Democracy from theory to success" In: European University Institute - Electronic Democracy Center, Sep. 25, 2008.

[22] Resultado da Votação do OP 2011. Disponível em <http://opdigital2011.pbh.gov.br/resultado-completo.php>. Acessado em 05/08/2013 às 20:58.

[23] Hull, D., West, G. and Cecez-Kecmanovic D. (2011) “Two Models of E-Democracy: A Case Study of Government Online Engagement with the Community”, In: World Academy of Science and Technology Conference, Paris, 2011.

[24] Larsson, A.O., Moe, H. (2011) “Studying political microblogging: Twitter users in the 2010 Swedish election campaign” In: New Media Society 2012 14: 729 - 21 November 2011.

[25] Lichtig. R. The History of Natural Language Processing Disponível em <http://www .ieeeghn.org/wiki/index.php/The_History_of_Natural _Language_Processing>. Acessado em 22 de julho de 2013 às 22:13.

[26] Carenini, M., Whyte, A., Bertorello, L. and Vanocchi, M. (2007) “Improving Communication in E-democracy Using Natural Language Processing” In: Intelligent Systems, IEEE vol: 22, Issue: 1, Jan.-Fev. 2007.

[27] Summary of D6.2 Final Tool Validation and E-Democracy Evaluation Report. Disponível em <http://itc.napier.ac.uk/itc/Documents/EDEN_D62_Summary.pdf>. Acessado em 2/11/2013 às 23:12.

[28] Rossi, C. Mídias Sociais: rumo à democracia participativa? Disponível em <http://sociol ogiacienciaevida.uol.com.br/ESSO/edicoes/37/artigo238948-1.asp>. Acessado em 28 de Novembro de 2013 às 16:32.

[29] Hussain, A.R.; Hammed, A. M; Hegde, N.P. (2011). "Mining Twitter using Cloud Computing". In: Information and Communication Technologies (WICT), p. 187 – 190, 11-14 Dez. 2011.

[30] Reips, Ulf-Dietrich, and Garaizar, Pablo. (2011). Mining Twitter: Microblogging as a source for psychological wisdom of the crowds. Behavior Research Methods, 43, 635-642.

[31] Priyanthan, P.; Ragavan, T.; Prasath, N.; Perera, A. (2012). "Opinion mining and sentiment analysis on a Twitter data stream". In: Advances in ICT for Emerging Regions (ICTer), p. 182 – 188, 12-15 Dez. 2012.

[32] Twitter apresenta o seu IPO, e ver o número dos usuários diários 100m, 500m e tweets por dia. Disponível em <http://infotecnomundo.com/2013/10/03/twitter-apresenta-o-seu-ipo-

e-ver-o-numero-dos-usuarios-diarios-100m-500m-e-tweets-por-dia/>. Acessado em 01/11/2013 às 12:58.

[33] Oussalah, M.; Challis, K.; Schnier, T. (2011). "A software system for data mining with Twitter". In: Cybernetic Intelligent Systems (CIS), 2011 IEEE 10th International Conference, p. 139 – 144, 1-2 Set. 2011.

[34] FAYYAD, U.M. et al. Advances in Knowledge Discovery and Data Mining. California: AAAI Press, 1996b.

[35] ADDRIANS, P. & ZANTINGE, D. Data Mining. Inglaterra: Addison-Wesley, 1996.

[36] Leal, L.M.N. Text Mining. Disponível em <http://www.dei.isep.ipp.pt/~paf/proj/Set2003/Text%20Mining.pdf> . Acessado em 01/11/2013 às 13:42.

[37] Leite, F.G. Text Mining: Extraindo conhecimentos utilizando essa tecnologia. Disponível em <http://pessoal.utfpr.edu.br/arildo/arquivos/Mono_FelipeLeite[1].pdf>. Acessado em 1/11/2013 às 14:12.

[38] Souza, T.; Lindgren, A. Mineração de Texto. Disponível em <http://rec.univercidade.edu.br/artigos/artigoE01_03.pdf>. Acessado em 2/11/2013 às 21:53.

[39] Morais E.A.M.; Ambrósio A.P.L. (2007). "Mineração de Textos". In: Technical Report – INF_005/07, Dez 2007.

[40] Rajaraman, A.; Ullman, J. D. (2011). "Data Mining". Mining of Massive Datasets. pp. 1–17.

[41] Jing, L.P.; Huang, H.K.; Shi, H.B. (2002). "Improved feature selection approach TFIDF in text mining". In: Machine Learning and Cybernetics, 2002. Proceedings. 2002 International Conference (Volume: 2), p. 944 – 946 vol 2.

[42] API Documentation. Disponível em <https://dev.twitter.com/start>. Acessado em 25 de Outubro de 2013 às 19:12.

[43] Twitter Firehose vs. Twitter api: what's the difference and why should you care?. Disponível em <http://www.brightplanet.com/2013/06/twitter-firehose-vs-twitter-api-whats-the-difference-and-why-should-you-care/>. Acessado em 2 de Dezembro de 2013 às 10:12.

[44] Câmara de SP aprova aumento do IPTU em primeira votação. Disponível em <http://www1.folha.uol.com.br/cotidiano/2013/10/1361728-camara-aprova-aumento-do-iptu-em-primeira-votacao.shtml>. Acessado em 2 de Dezembro de 2013 às 10:27.

[45] TJ-RJ suspende liminar e plano para professores volta a valer, diz Câmara. Disponível em <http://g1.globo.com/rio-de-janeiro/noticia/2013/10/tj-derruba-liminar-e-plano-de-cargos-de-professores-do-rio-volta-valer.html> Acessado em 2 de Dezembro de 2013 às 10:47.

[46] Com atraso, Via Binário é inaugurada no Rio e Perimetral fecha para testes. Disponível em <http://g1.globo.com/rio-de-janeiro/noticia/2013/10/com-atraso-binario-e-inaugurada-no-rio-e-perimetral-fecha-para-testes.html> Acessado em 2 de Dezembro de 2013 às 11:12.

[47] Projeto aprovado pela Alerj prevê devolução do IPVA em caso de roubo. Disponível em <http://g1.globo.com/rio-de-janeiro/noticia/2013/10/projeto-aprovado-pela-alerj-preve-devolucao-do-ipva-em-caso-de-roubo.html>. Acessado em 2 de Dezembro de 2013 às 11:31.

[48] SANEAMENTO É O BÁSICO! PELO FIM DOS PODERES ABSURDOS DA CEDAE!. Disponível em <http://www.meurio.org.br/na_atividade/21/assine_embaixo/saneamento-basico-para-todo-o-rio?utm_source=Mailee&utm_medium=email&utm_campaign=Saneamento+%C3%A9+B%C3%A1sico&utm_term=&utm_content=Saneamento+%C3%A9+B%C3%A1sico+pra+toda+a+lista+menos+MRDash+15/08>. Acessado em 2 de Dezembro de 2013 às 12:44.

[49] Construção de linha de metrô em Curitiba terá investimento de R$ 4,6 bilhões. Disponível em <http://piniweb.pini.com.br/construcao/Infraestrutura/construcao-de-linha-de-metro-em-curitiba-tera-investimento-de-300824-1.aspx>. Acessado em 1 de Novembro de 2013 às 15:37h.

[50] Governo vai liberar verba para obras de mobilidade. Disponível em <http://www.gazetadopovo.com.br/m/conteudo.phtml?id=1420042&tit=Governo-vai-liberar-verba-para-obras-de-mobilidade>. Acessado em 30 de Novembro de 2013 às 13:37h.

ANEXOS

ANEXO A

Lista de StopWords

A, à, agora, ainda, alguém, algum, alguma, algumas, alguns, ampla, amplas, amplo, amplos, ante, antes, ao, aos, após, aquela, aquelas, aquele, aqueles, aquilo, as, até, através, cada, coisa, coisas, com, como, contra, contudo, da, daquele, daqueles, das, de, dela, delas, dele, deles, depois, dessa, dessas, desse, desses, desta, destas, deste, deste, destes, deve, devem, devendo, dever, deverá, deverão, deveria, deveriam, devia, deviam, disse, disso, disto, dito, diz, dizem, do, dos, e, é, e', ela, elas, ele, eles, em, enquanto, entre, era, essa, essas, esse, esses, esta, está, estamos, estão, estas, estava, estavam, estávamos, este, estes, estou, eu, fazendo, fazer, feita, feitas, feito, feitos, foi, for, foram, fosse, fossem, grande, grandes, há, isso, isto, já, la, la, lá, lhe, lhes, lo, mas, me, mesma, mesmas, mesmo, mesmos, meu, meus, minha, minhas, muita, muitas, muito, muitos, na, não, nas, nem, nenhum, nessa, nessas, nesta, nestas, ninguém, no, nos, nós, nossa, nossas, nosso, nossos, num, numa, nunca, o, os, ou, outra, outras, outro, outros, para, pela, pelas, pelo, pelos, pequena, pequenas, pequeno, pequenos, per, perante, pode, pôde, podendo, poder, poderia, poderiam, podia, podiam, pois, por, porém, porque, posso, pouca, poucas, pouco, poucos, primeiro, primeiros, própria, próprias, próprio, próprios, quais, qual quando, quanto, quantos, que, quem, são, se, seja, sejam, sem, sempre, sendo, será, serão, seu, seus, si, sido, só, sob, sobre, sua, suas, talvez, também, tampouco, te, tem, tendo, tenha, ter, teu, teus, ti, tido, tinha, tinham, toda, todas, todavia, todo, todos, tu, tua, tuas, tudo, última, últimas, último, últimos, um, uma, umas, uns, vendo, ver, vez, vindo, vir, vos, vós.