

Análise de Sentimento no Contexto de Comentários a Projetos de Lei Relativos à Transparência

Roxana Lisette Quintanilla Portugal, Julio Cesar Sampaio do Prado Leite

Departamento de Informática, PUC-Rio, Rio de Janeiro, Brasil
{rportugal, julio}@inf.puc-rio.br

***Abstract.** In 2016, Engiel et al. found 10 bills highly related to transparency, using text mining driven by a transparency knowledge base, on the VotenaWeb site. In 2017, Portugal et al. text-mined comments on these 10 bills, in 164 of these comments, 6 bills were found to have a tendency or demand for transparency. The present work explores the comments from another perspective; we wanted to explore sentiment analysis to know the most negative comments by using the SentiStrength tool. As a result, of the 164 comments 60 were classified to be very negative, however, its accuracy is not a good one when dealing with the set of comments from citizens who approve the bills. We detailed a qualitative analysis of these comments with respect to the tool result.*

***Resumo.** Em 2016, Engiel et al. encontraram 10 projetos de lei, altamente relacionados à transparência, usando técnicas de mineração de texto e uma base de conhecimento sobre Transparência. Em 2017, Portugal et al. usaram os comentários feitos sobre os 10 projetos de lei; dos 164 comentários, 6 projetos foram identificados com uma tendência ou demanda por transparência. O trabalho de 2018 explora os comentários a partir de outra perspectiva, nós queremos saber quais são os comentários mais negativos usando análise de sentimentos, através da ferramenta SentiStrength. Como resultado, dos 164 comentários, 60 foram classificados como muito negativos. No entanto, seu desempenho não é bom quando se lida com o conjunto de comentários dos cidadãos que aprovam os projetos de lei. Uma análise qualitativa é apresentada para esse subconjunto de comentários.*

Palavras-Chave: Análise de Sentimento, Transparência, Projetos de Lei, Mineração de texto

1. Introdução

Existem vários desafios de como organizar grande volume de dados, sendo um deles a análise de comentários na mídia. A partir do ponto de vista da transparência, essas análises de comentários constituem uma oportunidade para procurar descobrir o que o cidadão está demandando.

Um *corpus* de comentários amplamente usado para a análise de sentimentos é o *Twitter*. O qual, pelas suas características de comentários pequenos e textos escritos de maneira informal, exige heurísticas mais elaboradas, além da classificação por polaridade, de comentários negativos ou positivos. No entanto, existe a possibilidade de

um texto ter ambos os sentimentos, assim como alguns serem mais positivos do que outros.

SentiStrength é um trabalho de Thelwall et al. (2012) que têm refinado suas heurísticas para o tratamento de textos curtos e informais. Um exemplo é o tratamento de textos mal escritos intencionalmente, por exemplo: a palavra “mal” expressa um sentimento menos negativo do que “maaal”, não obstante, em algoritmos que só determinam a polaridade, isto é considerado como uma falha na escrita e tem um procedimento que corrige a palavra.

Neste trabalho, verificamos os sentimentos expressos em comentários referentes a 8 projetos de lei, do sítio *VotenaWeb* (2009), relacionados à transparência. O *corpus* de 164 comentários foi classificado em três categorias: comentários de cidadãos que *aprovam*, *desaprovam*, ou que estão *indecisos* sobre um Projeto de Lei (PL). Os textos cumprem com as características para o qual o *SentiStrength* foi desenvolvido, textos curtos e informais. O principal desafio é verificar o desempenho do *SentiStrength* usando o léxico em Português que foi publicado pelos autores, no entanto, como um léxico não testado (*SentiStrength Tool*, 2018).

O trabalho está organizado da seguinte forma. A Seção 2 apresenta os comentários processados, contextualizando-os com a descrição de cada PL (Projeto de Lei). A Seção 3 discute as características da ferramenta utilizada. A Seção 4 relata resultados obtidos face à classificação manual por um dos autores. Concluímos resumizando os resultados, bem como indicando trabalhos futuros.

2. Projetos de Lei com indicadores de Transparência

Em 2016, Engiel et al. identificaram 10 PLs *+transparentes* por possuírem indicadores de transparência. Essa filtragem foi realizada utilizando técnicas de mineração de texto e conhecimento de transparência, representado em um catálogo SIG (*Softgoal Interdependence Graph*) Chung et al. (2000). O SIG de transparência do grupo ER-PUC Rio (2013) é uma modelagem de qualidades, que são utilizadas para a identificação de requisitos não funcionais (NFR) de transparência.

Desses 10 PLs, Portugal et al. (2017a, 2017b) descobriram que dois PLs, o PLC 370-2010 e PLV 8-2010 foram selecionados erroneamente.

Tabela 1. 10PL + transparentes (Engiel, 2016)

Projeto de Lei	% Voto (si)
pls-193-2013	97%
plc-1080-2011	96%
pls-79-2010	95%
plc-3497-2015	94%
pls-82-2009	93%
pls-387-2011	92%
plc-3305-2008	86%
pls-68-2010	72%
plc-370-2010	61%
plv-8-2010	48%

A Tabela 1, completa (com os projetos riscados), foi apresentada por Engiel et al. (2016) e mostrava a aprovação dos cidadãos sobre os 10 PLs + transparentes no portal *Votenaweb*. A Tabela 1 foi reordenada segundo o % de aprovação para facilitar a análise nas Subseções seguintes (2.1 a 2.8). Para cada PL de cada Subseção foi

consultada, na fonte¹ do trabalho prévio de Portugal et al. (2017b), quais palavras-chaves do catálogo de Transparência que filtraram esse PL, ou seja, o determinante para que a mineração de texto identificasse o referido PL como relacionado à transparência. Para cada Subseção, apresenta-se um trecho do PL, para efeito de contextualização. Para cada Subseção iremos analisar os comentários utilizando o software *SentiStrength* (2012) de maneira a entender como os comentários a cada PL da Tabela 1 se comporta sob a perspectiva de “sentimento”.

A análise do *SentiStrength* (2012) qualifica o sentimento de um texto tanto na escala: de 1 (positivo) até 5 (extremamente positivo), como na escala de -1 (negativo) até -5 (extremamente negativo). Se um texto tem a mesma qualificação, por exemplo (3, -3), é considerado um texto neutro. Para a análise a seguir, em cada Subseção, filtramos os comentários bem negativos, ou seja, aqueles que o *SentiStrength* qualifica de -3 até -5. Para cada Subseção, apresenta-se uma tabela com o comentário e as qualificações em ambas escalas (positiva e negativa) dada pela ferramenta.

2.1. PLS-193-2013

O PL tem por título: *Obrigar a OAB e os outros conselhos profissionais a se submeterem às regras da Lei de Acesso à Informação*. Tem 97% de aprovação dos cidadãos com 3392 votos. A situação atual do projeto é: *em tramitação*. O PL demanda a qualidade de *acessibilidade* (ver Fig.1). O PL tem 25 comentários, dos quais, após a análise de sentimentos, 6 foram classificados como bem negativos, ver Tabela 2.

Este projeto de lei obrigará a Ordem dos Advogados do Brasil – OAB e os outros conselhos de fiscalização de profissões regulamentadas (por exemplo: CRM, CRF, etc) a se submeterem às regras da Lei de Acesso à Informação. Ou seja, os conselhos de fiscalização não poderão recusar o fornecimento de informações quando for solicitado por alguém. Segundo o senador, os conselhos de fiscalização de profissões regulamentadas têm natureza de direito público e prestam serviço público, apesar de não fazerem parte da Administração Pública. É necessário garantir aos filiados destes conselhos o exercício da fiscalização de seus dirigentes, obtendo transparência das informações.

Figura 1. Palavras-chaves de transparência no PLS-193-2013

Tabela 2. Análise de Sentimentos nos comentários do PLS-193-2013

Comentários	+	-
Não acho que eles possam se eximir, mas apertar o pescoço dos Conselhos é essencial. Da OAB então, nem se fala.	1	-3
Já é assim, não é?:	1	-3
@Kelly e @Raphael, porque a OAB, efetivamente, é um órgão que governa a vida dos cidadãos. Não é governamental, de fato, mas todo advogado é obrigado a ser membro e passar no exame deles, de forma que funciona como uma agência governamental	2	-3
Os conselhos profissionais não tem verba pública envolvida então a lei de acesso à informação não se aplica, se aplica apenas aos profissionais de cada área registrados, estes sim podem solicitar acesso à informações, o público em geral não tem porque ter acesso às informações, porque um administrador ou um engenheiro precisa saber das contas da OAB? se quem contribui são os advogados? Os Advogados sim precisam ter acesso a informações da OAB, engenheiros do CREA e assim por diante.	1	-3
Porque a OAB tem que se submeter a lei de acesso a informação, se não recebe nenhum subvenção do governo? A OAB sobrevive das anuidades dos advogados, e não recebe nenhum centavo do governo. Portanto, quem tem que pedir a prestação de contas são os próprios advogados. Lei de acesso à informação é para conselhos de classe que recebem do governo não para a oab.	1	-3
OAB não é órgão do Estado...Instituição anômala!	1	-3

¹ <https://github.com/nitanilla/Mining-Brazilian-bills-related-to-Transparency/issues> [issues 12, 13]

2.2 PLC-1080-2011

O PL tem por título: *Obrigará todos os senadores, deputados e vereadores a realizar audiências públicas mensal com o objetivo de prestar informações sobre sua atuação parlamentar*, e tem 96% de aprovação dos cidadãos com 1735 votos. A situação atual do PL é: *em tramitação*. A Figura 2 aponta as características de transparência do PL o qual foi classificado no grupo de PL's que demandam *informatividade*. Este projeto contém 67 comentários. Tabela 3 apresenta os 12 comentários que foram classificados como bem negativos.

assinaturas em caso de audiência pública por meio presencial. Senado Federal, a Câmara dos Deputados, as Assembléias Legislativas e as Câmaras de Vereadores deverão disponibilizar em seus sites oficiais na Internet espaço destinado à consulta para todas as pessoas interessadas contendo informações sobre as audiências públicas realizadas e as que já estejam designadas pelo Parlamentares que as integram, individualmente. As Casa legislativas deverão divulgar uma listagem com os nomes dos Parlamentares que não realizaram as audiências públicas obrigatórias. As audiências públicas equivalem a dias normais de expediente. O parlamentar que não realizar a audiência terá descontos diretos em seu salário na mesma proporção do dia não trabalhado. Segundo o deputado, é direito de toda a sociedade pedir que o agente público preste informações de sua atuação, portanto, o objetivo deste projeto de lei é implantar mais um canal de informação e transparência junto a todo cidadão, que terá o direito de participar com maior afinidade dos mandatos parlamentares que o representa.

Figura 2. Palavras chaves de transparência no PLC-1080-2011

Tabela 3. Análise de Sentimentos nos comentários do PLC-1080-2011

Comentários	+	-
O eleitor merece um follow up, para confirmar ou não se seu candidato está correspondendo a sua expectativa.	1	-4
Pela lógica deveria prestar contas do seu trabalho, mas já que isso não acontece seria válido esse projeto.	1	-3
Embora falte lapidar este projeto, pois poderiam agilizar o processo de informação, acredito que será um passo para diminuirmos o maior câncer deste, ou seja, a Corrupção!!!	1	-4
Isso poderia ser espontâneo, mas já que não é...	1	-3
Eles vão fazer as audiências e ainda vão enrolar a gente se a gente deixar. Não adianta o político prestar contas se o povo não presta atenção ;)	1	-3
voto sim com a pulga atrás da orelha, sabendo que a hora-extra das excelências vale quanto NÃO pesa e que estão doidos para faturar planilhas-extras de horas-extras através de assessores-extras, ghostwriters-extras, secretários-extras etcoeteras-extras	1	-3
Historinha dele, pois os pelegos vão para a audiência e os canalhas vão aprovar os projetos que os pelegos apoiaram! Vindo da esquerda, eu não confio!	1	-3
Concordo que seja na transparência pela web mesmo. Onde não haveria custo e todos nós poderíamos acompanhar. Eles tem que fazer o trabalhos deles. Concordo que se votar sim, seria mais uma perda de tempo e aumento dos gastos públicos.	1	-3
Eles têm assessores para auxiliá-los. E muitos! Não se perderia tempo; seria tempo investido. E não seria muito tempo.	1	-3
Com o voto distrital o representante seria de fato representantes, porque atualmente não é isso que ocorre, simplesmente votam aquilo que o partido/executivo manda e a população que se dane. Exemplo: recomposição das aposentadorias do INSS. Governo abocanhou o dinheiros do INSS e agora está fazendo o mesmo com o FGTS.	1	-3
Não sou político, mas acho desnecessário tais audiências. Se alguém tem tempo para isso e deseja saber o desempenho, propostas e feitos de parlamentares basta comparecer às sessões nas casas legislativas, ou assistam as TVs. Câmara e Senado. Seria mais um gasto às custas do erário público.	1	-3
Prestar contas é obrigação dos políticos. Mas Audiência Pública pra cada um vai ser um gasto público desnecessário, ainda mais para um público que não acompanha.	1	-3

2.3. PLS-79-2010

O PL com título: *Determinará obrigações para a realização de audiências públicas*, tem 91% de aprovação dos usuários com 136 votos. O estado atual do PL é: *em tramitação*. O PL demanda duas qualidades: *acessibilidade* e *informatividade* (ver

Fig.3). Em tabela 4 se mostra o único comentário de este projeto: *quais obrigações?*. O comentário foi classificado como neutro.

atingida pela política. A audiência deverá ter lista de identificação de todos os presentes que comporão a ata final. A verificação sobre quais são os dois jornais de maior circulação fica por conta do IVC (Instituto Verificador de Circulação). São direitos de qualquer cidadão interessado na política pública: opor-se ao edital de convocação, apontando motivos, até dois dias após a publicação; manifestar-se oralmente durante a discussão (dependendo da disponibilidade e tempo); solicitar ao Ministério Público uma nova audiência, com fundamentos e motivos (se não responder em até 2 dias, será considerada negada); solicitar, sempre que possível, a realização de plebiscito na região (ocorrendo somente após autorização legislativa específica); ter acesso a qualquer informação sobre a política a ser debatida, podendo pedir essas informações a qualquer órgão; Não se tratando de uma política delimitada geograficamente, deverá se consultar a população dos municípios envolvidos. Os municípios com mais de 100.000 habitantes deverão disponibilizar mecanismos de consulta às audiências públicas pela internet. Segundo o senador, as audiências garantem uma boa participação popular, mas pode ser algo manipulado, pois não existe um marco legal que garanta os direitos do cidadão participante. Assim, o projeto pretende normatizar as audiências públicas, focando na publicidade e transparência. Raimundo Colombo afirma que "O Brasil de hoje não mais compactua com políticas públicas decididas em gabinetes de burocratas ou com discussões que não envolvem os verdadeiros beneficiários das políticas públicas."

Figura 3. Palavras chaves de transparência no PLS-79-2010

Tabela 4. Análise de Sentimentos no comentário do PLS-79-2010

Comentários	+	-
Quais obrigações?	1	-1

2.4. PLC-3497-2015

O PL tem por título: *Tornará obrigatória a divulgação, na internet, da folha de pagamento dos funcionários da Administração Pública Direta e Indireta, incluindo adicionais, auxílios e gratificações*, e tem 94% de aprovação dos cidadãos com 2372 votos. A situação atual do PL é: *em tramitação*. A Fig. 4 mostra que o PL demanda principalmente *acessibilidade*. Este projeto contém 20 comentários, dentre os quais 6 comentários são bem negativos (ver Tabela 5).

Este projeto de lei vai fazer uma alteração na Lei de Acesso à Informação que obriga os órgãos públicos a divulgarem informações de interesse coletivo. Esta projeto vai obrigar que os órgãos e entidades públicas deverão divulgar em local de fácil acesso, os registros das despesas, inclusive com pagamento de servidores e empregados, englobando na remuneração publicada os vencimentos ou subsídios, somados à adicionais, auxílios e gratificações de qualquer natureza. Segundo o deputado, atualmente não são todos os órgãos que divulgam a aplicação dos seus recursos, principalmente aqueles que fazem parte da Administração Indireta. Para o deputado, é importante garantir e cobrar mais transparência no controle com os gastos do erário e fiscalização do dinheiro público.

Figura 4. Palavras chaves de transparência no PLS-3497-2015

Tabela 5. Análise de Sentimentos nos comentários do PLS-3497-2015

Comentários	+	-
VÃO LAMBER SABÃO, POLÍTICOS!!!. PAREM DE FERRAR COM A VIDA DAS EMPRESAS QUE DÃO EMPREGO PARA A POPULAÇÃO!!!	2	-3
As empresas privadas vão divulgar também? Meu vizinho não precisa saber quanto recebo...	1	-3
Já existe isso, rapá... para de jogar a culpa da robalheira dos políticos em cima dos coitados que fazem concurso pra trabalhar...	3	-4
Por quê não inclui a folha de pagamento dos políticos, inclusive com adicionais, extras, passagens, hospedagens, auxílios.....	1	-3
Já não é obrigatório?	1	-3
Deveria obrigar a divulgação do pagamento dos não concursados e o grau de parentesco destes com administradores, políticos, juízes e ministros.	1	-3

2.5. PLS-82-2009

O PL com título: *obrigará as empresas estatais federais a divulgarem, de 3 em 3 meses, as despesas realizadas com publicidade e patrocínio no Diário Oficial da União e em*

jornais de circulação nacional, para aumentar a transparência dos gastos públicos, tem 94% de aprovação dos cidadãos com 161 votos. A situação atual do PL é: em tramitação. A Fig.5 mostra que o PL demanda principalmente publicidade. Este PL contém 3 comentários, e o SentiStrength qualificou apenas um como negativo. No entanto não seria qualificado com bem negativo por ter apenas -2 (ver Tabela 6.)

nacional. As despesas deverão ser consolidadas e divulgadas de 3 em 3 meses. A divulgação das despesas deverá separar os gastos realizados com publicidade dos gastos realizados com patrocínio, devendo especificar: - a entidade contratada ou patrocinada; - o nome e o número de cadastro da pessoa física ou jurídica contratada ou patrocinada; - o número de identificação e o valor de cada contrato ou patrocínio; - a data de celebração e o período de vigência de cada contrato ou patrocínio; - informação sobre a existência de licitação prévia e, nos casos em que não houver licitação, é necessário justificar a dispensa ou a inexigibilidade da licitação; - breve relato da finalidade do patrocínio ou contrato de publicidade; - o valor total dos contratos e dos patrocínios no trimestre, valor acumulado do exercício vigente e no exercício financeiro anterior. O valor total das despesas com publicidade e patrocínio de cada empresa estatal federal, em cada exercício financeiro, não poderá ser superior a 15% do valor médio das respectivas despesas dos quatro exercícios financeiros anteriores. O descumprimento desta lei sujeitará a autoridade máxima da entidade às penas previstas para o crime de responsabilidade. Segundo o senador, este projeto de lei tem a finalidade de aperfeiçoar a transparência e o controle das despesas realizadas pelas empresas estatais federais com publicidade e patrocínio.

Figura 5. Palavras chaves de transparência no PLS-82-2009

Tabela 6. Análise de Sentimentos nos comentários do PLS-82-2009

Comentários	+	-
Só na internet já seria o suficiente. Teria um gasto desnecessário de dinheiro indo para os jornais	1	-2

2.6. PLS-387-2011

O PL com título: *Obrigará as universidades públicas e as unidades de pesquisa a criarem arquivos online, de acesso livre, para guardar toda a produção científica dos estudantes, professores, pesquisadores e colaboradores*, tem 91% de aprovação dos cidadãos com 608 votos. A situação atual do PL é em tramitação. Em Fig. 6 mostra-se que o PL demanda principalmente acessibilidade. Dos 15 comentários deste projeto, 2 foram qualificados como bem negativos (ver Tabela 7).

(dados sobre os dados da produção científica); - informações que descrevam a produção técnico-científica, a bibliografia, informações relacionadas com as questões de direitos, mantendo-os em acesso restrito enquanto durar a restrição, devendo disponibilizar o acesso ao seu completo teor a partir do momento que acabar a restrição. O depósito deverá oferecer também a possibilidade de o usuário, que tenha interesse em acessar um documento de acesso restrito, solicitar uma cópia deste documento diretamente ao pesquisador por e-mail. As agências de pesquisa e universidades deverão incluir em suas memórias de cálculo o número de artigos publicados em revistas com revisão por pares que foram depositados em depósitos institucionais, para avaliar a produção científica do pesquisador. Será criado o comitê de alto nível, que será constituído por representantes dos principais segmentos da comunidade científica envolvidos da produção da pesquisa científica, com o objetivo de propor uma política nacional de acesso livre à informação científica, além de estimular professores, pesquisadores e colaboradores a obedecer esta lei, propondo ações e medida que promovam o fluxo da informação científica. Segundo o senador esta medida dará mais visibilidade e transparência aos trabalhos produzidos pelas universidades, promovendo o conhecimento científico e cultural. Ainda de acordo com o deputado, os depósitos institucionais farão com que o governo invista mais em ciência e tecnologia, além de incentivar na elaboração de políticas de promoção de ciência e tecnologia para o país.

Figura 6. Palavras chaves de transparência no PLS-387-2011

Tabela 7. Análise de Sentimentos nos comentários do PLS-387-2011

Comentários	+	-
Não de acesso livre. Há estudos que envolvem a soberania nacional e não devem ficar expostos a outros países.	1	-3
Voto não, pois existem pesquisas confidenciais.	1	-3

2.7. PLC-3305-2008

O PL com título: *Regulamentará a licitação e a contratação de serviços de publicidade por parte da administração pública*, tem 84% de aprovação com 105 votos. A situação

atual do PL é em tramitação. O PL foi classificado principalmente no grupo de PL's que demandam *auditabilidade* (ver Fig. 7). Este PL contém 3 comentários, dois quais 2 foram classificados como bem negativos.

passarão a ter de ser licitados por meio de procedimentos autônomos. Para evitar a condução imoral de procedimentos licitatórios, propomos nesse projeto a redefinição da formação das comissões responsáveis pelo processamento e julgamento destas licitações, que obrigatoriamente terão de ser compostas por pessoas escolhidas por sorteio, incluindo-se especialistas na matéria que não mantenham qualquer vínculo funcional ou contratual com o Poder Público. O projeto busca, ainda, alterar o processamento da licitação de serviços de publicidade, fazendo com que, durante o julgamento da proposta técnica sejam desconhecidos, dos membros da comissão de licitação, os autores das propostas que serão julgadas. Buscará alterar radicalmente os mecanismos de controle da execução desses contratos. Passarão a ser exigidos o cadastramento de fornecedores das empresas de publicidade contratadas, a realização de orçamentos prévios, a disponibilização de dados da execução dos contratos pela rede mundial de computadores, para assegurar-se a transparência e a possibilidade de ampla fiscalização pela sociedade, e a satisfação maior da moralidade administrativa.

Figura 7. Palavras chaves de transparência no PLC-3305-2008

Tabela 8. Análise de Sentimentos nos comentários do PLC-3305-2008

Comentários	+	-
Administração pública, seja qual for, inclusive empresas estatais não deveriam fazer propaganda. Se for informação de interesse público, os jornais, escritos, falados, televisivos fazem sem cobrar nada. O dinheiro é do povo e não deve ser gasto com propagandas para enaltecer o partido que estiver no poder, pois ao final das contas essas propagandas do governo só servem para isso.	2	-3
a regulamentação tem que ser na representação do povo... e não de empresas!	1	-3

2.8 PLC-68-2010

O PL tem por título: *Modificará o sistema de arquivamento dos votos em urnas eletrônicas e os mecanismos de transparência dos votos*, e tem 72% de aprovação dos cidadãos com 184 votos. A situação atual do PL é: *em tramitação*. O PL em Fig. 8 demanda *acessibilidade*. Dos 21 comentários do PL, 11 foram classificados como bem negativos (ver Tabela 9).

TSE poderá auxiliar no desenvolvimento de pesquisas, junto à órgãos de pesquisa e com a comunidade acadêmica para auxiliar na melhoria contínua da segurança e da transparência do sistema eletrônico de votação. A Justiça Eleitoral fica autorizada a convocar, sob pena de cancelar as inscrições, os eleitores para recadastramento ou atualização de dados (incluindo os de dados biométricos ou fotos) desde que observadas as orientações específicas do TSE. Essa convocação deverá ter uma ampla divulgação, para orientar o eleitor sobre os locais e horários que ele deverá se apresentar. Os dados do cadastro eleitoral serão administrados e utilizados exclusivamente pela Justiça Federal, e o TSE deverá regulamentar as hipóteses e formas de acesso a esses dados. As hipóteses só poderão ser relativas aos dados de nome, filiação, data de nascimento, sexo, dados biométricos, fotografia e número da inscrição eleitoral. Para fazer o cadastro biométrico, a Justiça Eleitoral poderá requisitar servidores (federais, estaduais e municipais), e também instalações e serviços de órgãos da União, dos Estados, do Distrito Federal e Municípios. O TSE vai baixar as instruções dessa lei, para definir o cronograma de implantação do cadastro biométrico, os procedimentos no caso de duplicidades verificadas pelos dados biométricos; a forma da troca de dados; os procedimentos e a forma de verificação da biometria, para habilitar os eleitores aos votos; e os procedimentos de auditoria e segurança, envolvidos nas urnas com a verificação biométrica da identidade do eleitor. Segundo Azeredo, a justiça eleitoral tentou utilizar um sistema de impressão dos votos, que iam para um local lacrado, mas isso gerava mais problemas que soluções, pois aumentava a fila nas eleições, aumentava o número de votos brancos e nulos, e o número de eleitores que tinham dificuldade para finalizar seu voto. Além disso, o sistema não apresentava segurança e transparência adequada. Dessa forma, o projeto tem por objetivo evitar fraudes eleitorais, como os votos duplos (a pessoa vota duas vezes em localidades diferentes), aumentando a segurança e transparência das eleições.

Figura 8. Palavras chaves de transparência no PLC-68-2010

Tabela 9. Análise de Sentimentos nos comentários do PLC-68-2010

Comentários	+	-
Imagina que absurdo! Só no Brasil, com uma população alheia a política, que se permite que uma lei assim venha a ser proposta. Se esse projeto for aprovado, eles controlarão quem vai ser "eleito" e pronto, senão, pode nos forçar a votar em determinado candidato, se os nossos dados ficarem vinculados ao voto, como já é suspeita essa prática lá na Venezuela, onde o governo faz chantagem com o cidadão, prejudicando sua carreira em função do voto de cabresto.	2	-3
Absurda a pretensão do Senador. Compactuar com isso é no mínimo ignorância. Sem conferenciado voto, não há democracia nem garantia. É o primeiro passo para o fim dos direitos civis. Sem transparência, não temos eleição!	1	-3
São falsos os argumentos do Sen. Azeredo. Ele pretende novamente substituir o voto impresso CONFERÍVEL pelo eleitor pelo seu voto virtual INCONFERÍVEL pelo eleitor. O voto virtual que o Azeredo quer recriar elimina totalmente a transparência da apuração, POIS O ELEITOR NÃO TERÁ COMO CONFERIR se seu voto foi gravado como queria. É um procedimento que não atende ao Princípio da Independência do Software em sistemas eleitorais que está sendo adotado em todo o mundo.	2	-3
Azeredo, uma vergonha para nossa democracia, além de Coordenador do Mensalão Mineiro, é agora criminoso eleitoral, todos os países podem recontar os votos menos o Brasil. O senador tem alguma coisa contra a transparência?	1	-3
Mais uma peraltice do Sen. Azeredo, que já devia estar gozando de merecida aposentadoria, em lugar de se prestar ao triste papel de laranja do TSE. Toda vez que ele se mete a propor leis sobre o que não entende, é um desastre. Pois de segurança de dados ele entende menos do que eu, engenheiro eletrônico,	1	-4

entendo de cirurgia cardíaca, e tenta enganar aos incautos, vendendo seu mal intencionado projeto com justificativas estapafúrdias, assopradas sabe-se lá' por quem... Ele quer que continue valendo o lema EU SEI EM QUEM VOTEI. ELES TAMBEM. MAS SO ELES SABEM QUEM RECEBEU O MEU VOTO...		
Mais uma do velho conhecido Eduardo Azeredo. Ele novamente quer acabar com a possibilidade de recontagem dos votos. Para esse senador, projetos como ficha limpa e voto impresso são um problema. Políticos ligados ao Mensalão e à corrupção não querem a transparência e a segurança do processo eletrônico de votação. Por isso voto não a esse projeto anti- democrático.	2	-3
O voto eletrônico é completamente falho. Precisamos de sermos respeitados nas nossas escolhas. Não é atoa que os outros países não querem adotar a urna eletrônica e as que adotaram, devolveram.	1	-3
O sistema de voto eletrônico é totalmente falho e não temos segurança alguma nos nossos votos. Não sabemos se o nosso voto vai mesmo pro candidato que escolhemos. E o senhor senador sabe bem disso. Acho que deveria apoiar a ética e a liberdade das pessoas votarem e que sejam respeitados as nossas escolhas..... os nossos votos.	2	-3
O senador define uma solução técnica como projeto de lei. O que ele deveria definir no projeto é o resultado prático, não a solução técnica, a qual tem de ser estudada por um equipe técnica qualificada para atender ao requisito prático.	2	-3
Não foi esse o que renunciou o cargo de deputado, acusado de mensalão do psdb em minas? sem noção essa PL.	1	-3
Ainda considero o sistema brasileiro falho por conta de não ter sistema de checagem amostral, via impressão dos votos q deveriam ser depositados em urnas, como é feito em qq país democrático. Causa estranheza um sistema de votação tão simples e célere não ser adotado por outros países.	2	-3

3. Discussão

A análise de sentimentos, especificamente para a língua portuguesa, tem sido abordada por trabalhos como o *SentiMeter-Br* (2013), o qual apresenta melhores resultados que o *SentiStrength* (2012), no entanto o *SentiMeter-Br* foi desenvolvido para o domínio de Cosméticos. Outro trabalho, *PIRPO* (2012) é similar aos dois primeiros por utilizar a estratégia baseada em um léxico, contudo o léxico de *PIRPO* foi realizado para o domínio de Hotéis. Ao melhor do nosso conhecimento o *SentiStrength* é o único a ter sido validado em várias bases, isto é, comentários de Myspace, Last FM, YouTube, e Flickr entre outros, para extrair um léxico suficientemente robusto para ser aplicado em outros contextos sociais. Aspectos como o tratamento de um corretor ortográfico que não corrige palavras que provavelmente estão escritas de maneira incorreta intencionalmente, um dicionário de *emoticons*, e o tratamento de palavras que dão uma maior ênfase (*strength*) fazem com que o *SentiStrength* seja apropriado para os textos como os comentários de *Votenaweb* (2009).

Um dos pontos fracos de *SentiStrength* é o léxico existente em Português, o qual não foi testado. Não se encontrou evidências do processamento de dito léxico. Um bom processamento de léxico, como indicado por Thelwall, (2017), demoraria pelo menos um mês para traduzir para o Português o léxico existente na língua inglesa, a criação de um *corpus* de 1000 textos previamente classificados por humanos, e o refinamento do léxico com opções da ferramenta através do exame de classificações incorretas no *corpus*. Por fim, a acurácia desse classificador deveria ser testada em um segundo *corpus* de textos. Existem outras limitações da ferramenta, mas, no momento, essa nos parece a mais importante para que se possa equiparar seu desempenho com o existente na língua Inglesa.

4. Resultados

Para verificar os resultados da análise de sentimentos apresentados na Seção 3, um dos autores fez a verificação manual dos comentários qualificados como bem negativos (de -3 até -5). A Tabela 10 mostra a quantidade de acertos do *SentiStrength*. Se mostra

também o grupo ao qual os comentários pertencem segundo a aprovação dos cidadãos. *Sim* para PL aprovados, *Não* para PL não aprovados, e *Indeciso*.

Tabela 10. Acurácia do *SentiStrength* em classificar comentários negativos (de -3 a -5)

	Número de comentários de cidadãos que votaram (sim/não/indeciso) nos PLs		
	Sim	Indeciso	Não
SentiStrength	18	13	29
Humano	5	10	22
	27%	76%	75%

Em Tabela 11 mostramos as palavras do léxico de *SentiStrength* que estão nos comentários que foram corretamente classificados na tabela 10. Mostra-se também a categorização de léxico que existe na ferramenta.

Tabela 11. Palavras que classificaram corretamente os comentários negativos

SIM	EmotionLookupTable	Vão [-2], não [-2], superficial [-2], muito[3]
	BoosterWordList	Muito[1]
NÃO	EmotionLookupTable	Vão [-2], não [-2], perda[-2], propaganda[-1], deve[-2], gasto[-2], culpa[-4], absurdo[-2], chantagem[-3], prejudicando[-3], pretensão[-2], vergonha[-3], contra[-2], Azarado[-2], triste[-4], desastre[-2], enganar[-2], mal[-3], velho[-2] problema[-2]
	BoosterWordList	deve[-1]
	Rule	!!! [+0.6 EmphasisInPunctuation]
INDECISO	EmotionLookupTable	não [-2], gasto[-2], simples[-2], Caro[-2], dúvida[-2]

Sobre os falsos positivos, verificou-se que vários comentários foram classificados como negativos por conter apenas a palavra **não** [-2]. Temos outros casos onde a ferramenta falhou:

Comentário no PLC-1080-2011 que tem -4 e não é negativo:

Embora falte lapidar este projeto, pois poderiam agilizar o processo de informação, acredito que será um passo para diminuirmos o maior câncer [-3] deste, ou seja, a Corrupção [-1]!!!

Comentário no PLC-68-2010 que por ter escrito Azarado em vez de Azeredo foi mais negativo:

Mais uma peraltice do Sen. Azarado, que já devia estar gozando de merecida aposentadoria, em lugar de se prestar ao triste papel de laranja do TSE...

Outros comentários têm palavras informais do português que não estão no léxico: *enrolar*, *xingarem*, *ferrar*, *roubalheira*, *rapaz*. As duas últimas se acharam mal escritas nos comentários como: *robalheira* e *rapá*. Outro exemplo é a palavra *apenas* que não se encontra no léxico, mas aparece significativamente na amostra utilizada.

5. Conclusão

Este trabalho teve como objetivo verificar a eficácia de *SentiStrength* aplicado ao contexto de Projetos de Lei relacionados à Transparência. Os resultados têm boa acurácia nos grupos de comentários *negativos* e *indecisos*, no entanto no grupo de comentários *positivos* (a favor do PL) o *SentiStrength* qualifica como negativo vários comentários por conter apenas a palavra *não*. Como trabalhos futuros, vemos que é possível realizar a estratégia proposta por Thelwall (2017) para treinamento no contexto específico, dado que os comentários no *VotenaWeb* já estão classificados (voto sim, não, indeciso) com o qual é possível recuperar os 1000 itens recomendados para

treinamento. Outro trabalho é evoluir o léxico em Português utilizando o conceito do LAL (Léxico Ampliado da Linguagem) (Engiel et al., 2015).

Segundo as informações de *VotenaWeb* (2009)², nenhum dos 8 PLs, com vários indicadores de demanda de transparência, já se transformou em Lei. Acreditamos que a estratégia apresentada poderia ajudar a filtrar os comentários dos cidadãos, para que os congressistas possam se apropriar da demanda que está sendo expressa em redes sociais, no caso tomando o *Votenaweb* como fonte.

Referências

- Engiel, P., Portugal, R.L.Q., do Prado Leite, J. C. S. Descobrimos projetos de lei relacionados a transparência. IV Workshop de Transparência em Sistemas, Rio de Janeiro - RJ, Brasil. (2016)
- Portugal, R. L. Q., Engiel, P., & do Prado Leite, J. C. S. (2017a). Existe uma Demanda de Transparência? Análise de comentários à Projetos de Lei. In V Workshop de Transparência em Sistemas. São Paulo, Brasil.
- Thelwall, M., Buckley, K., & Paltoglou, G. (2012). Sentiment strength detection for the social web. *Journal of the Association for Information Science and Technology*, 63(1), 163-173.
- SentiStrength Tool, 2018. Em (<http://sentistrength.wlv.ac.uk>). Acessado em 02/04/2018.
- Chung, L., & do Prado Leite, J. C. S. (2009). On non-functional requirements in software engineering. In *Conceptual modeling: Foundations and applications* (pp. 363-379). Springer, Berlin, Heidelberg.
- ER – PUC-Rio, 2013. Disponível em (http://transparencia.inf.puc-rio.br/wiki/index.php/Catálogo_Transparência). Acessado em 20/03/2017.
- Portugal, R. L. Q., Engiel, P., Roque, H., & do Prado Leite, J. C. S. (2017b). Is There a Demand of Software Transparency? In *Proceedings of the 31st Brazilian Symposium on Software Engineering* (pp. 204-213). ACM.
- Votenaweb, 2009. Disponível em (<http://www.votenaweb.com.br/>). Ac. Em 02/04/2018
- Thelwall, M. (2017). The Heart and soul of the web? Sentiment strength detection in the social web with SentiStrength. In *Cyberemotions* (pp. 119-134). Springer, Cham.
- Rosa, R. L., Rodriguez, D. Z., & Bressan, G. (2013). SentiMeter-Br: A new social web analysis metric to discover consumers' sentiment. In *Consumer Electronics (ISCE), 2013 IEEE 17th International Symposium on* (pp. 153-154). IEEE.
- Chaves, M. S., de Freitas, L. A., Souza, M., & Vieira, R. (2012). Pirpo: An algorithm to deal with polarity in portuguese online reviews from the accommodation sector. In *International Conference on Application of Natural Language to Information Systems* (pp. 296-301). Springer, Berlin, Heidelberg.
- Engiel, P., Pivatelli, J., Moura, P., Portugal, R. L. Q., & do Prado Leite, J. C. S. (2015). Um processo colaborativo para a construção de léxicos: o caso da divulgação de transparência. In *WER*.

² Em 02/04/2018