

COPPER - A Web-based platform for Characterization and Outlining of Popular Participation for Enhancing Regulation

Adriano Pereira, Wagner Meira Jr., Israel Guerra, Michel Boaventura, João Victor Bárbara, Geraldo Franciscani, Matheus Gonçalves, Pedro Nascimento
DCC UFMG

Belo Horizonte, Minas Gerais 31270-901
{adrianoc,meira,guerra,michel,jvictor,gfranciscani,souzagonc,l.pedro.nascimento}@dcc.ufmg.br

Newton Calegari, Diogo Cortiz, Vagner Diniz
CEWEB.br - NIC.br

CGI.br - Comitê Gestor da Internet Brasileira
Sao Paulo, SP
{newton,diogo,vagner}@nic.br

ABSTRACT

The growth and popularization of the Web has enabled several solutions of what has called e-democracy or cyberdemocracy. The most frequent manifestations of cyberdemocracy occur in the interaction between the political system and citizens, in particular mechanisms such as direct participation and debates, which tend to facilitate and amplify citizen participation in view of the direct interaction that facilitates and the removal of Geographical limits for participation. This work presents the proposal, implementation and evaluation of an analytical intelligence platform for the social participation process, which contemplates all the traditional steps of the process of knowledge discovery in databases. The tool was validated using actual data from a public consultation about Personal Data Protection.

KEYWORDS

Popular Participation, Web 2.0, e-Gov, Data Mining, Data Science

1 INTRODUÇÃO

O crescimento e capilarização da Web tornou possível e tem viabilizado diversas soluções do que se tem denominado e-democracia ou ciberdemocracia [2]. Sem perda de generalidade podemos definir ciberdemocracia como a prática política realizada através de tecnologia digital. Há vários desafios na efetiva implementação da ciberdemocracia como participação, transparência, preservação de direitos e liberdades, dentre outros [3, 4].

As manifestações mais frequentes da ciberdemocracia ocorrem na interação entre o sistema político e os cidadãos, em particular por mecanismos como a participação direta e debates, os quais tendem a facilitar e amplificar a participação dos cidadãos tendo em vista interação direta que propicia e a remoção de limites geográficos para a participação [1]. Desta forma, é possível materializar a ciberdemocracia em todo o processo cidadão: avaliação das necessidades, obtenção da informação, decisão, avaliação e aperfeiçoamento das ações. Por outro lado, assim como em vários cenários de aplicação habilitados pela internet, há o risco de que desigualdades e outros

formas de injustiça sejam acentuadas. Mais ainda, o uso massivo de tecnologias traz outras questões como a necessidade de analisar, caracterizar, interpretar, qualificar e quantificar as participações digitais daqueles cidadãos.

Um aspecto do processo de ciberdemocracia que é objeto deste artigo é a construção da legislação, uma vez que o processo de participação popular se mostra uma estratégia cada vez mais frequente e efetiva de construção de legislação da população como um todo. No Brasil já existe toda uma legislação sobre o processo em si e o mesmo se encontra em constante evolução e aperfeiçoamento, em particular com relação aos mecanismos de participação e verificação de legitimidade da mesma. Uma etapa fundamental que sucede a participação em si é a análise e o entendimento das opiniões expressas na participação, tarefa cujo desafio é proporcional ao número de participantes, sua intensidade de participação e a complexidade do tópico da consulta.

Dada a importância não só do tema, mas a relevância da construção de metodologias de análises de participação social na construção legislativa que se faz cada vez mais presente na construção de diversas políticas públicas, há a necessidade de se pensar como rotular, validar e incorporar todas as informações de forma objetiva e analítica. A plataforma Copper, objeto deste trabalho, se insere neste contexto, a construção de uma plataforma de inteligência analítica que auxilie no processo de análise e entendimento da participação popular em consultas públicas.

Há uma grande variedade de plataformas e ferramentas de inteligência analítica, e vamos brevemente discutir algumas, contextualizando a pertinência da plataforma COPPER. Podemos organizar a nossa discussão em três dimensões: técnicas de inteligência analítica, metáforas visuais e nível de abstração. Em termos de técnicas de inteligência analítica, já há um grande número de implementações disponíveis, as quais podem ser integradas de variadas formas. Podemos destacar Python Scikit-Learn¹ e R² como dois exemplos de bibliotecas de técnicas de grande cobertura. Em termos de metáforas visuais, ressaltamos D3³, como sendo uma plataforma próxima do estado da arte em termos de visualização. Entretanto, ainda há uma carência de metáforas visuais para modelos e padrões gerados por técnicas de inteligência analítica. Finalmente, há

In: XVI Workshop de Ferramentas e Aplicações (WFA 2017), Gramado, Brasil. Anais do XXIII Simpósio Brasileiro de Sistemas Multimídia e Web: Workshops e Pôsteres. Porto Alegre: Sociedade Brasileira de Computação, 2017.

© 2017 SBC – Sociedade Brasileira de Computação.

ISBN 978-85-7669-380-2.

¹scikit-learn.org/

²cran.r-project.org/

³d3js.org/

diversas plataformas de caráter geral, como Knime ⁴ e Orange ⁵, que não tem foco em um problema específico. Desta forma, até onde pudemos apurar, não há uma plataforma que se assemelhe à COPPER, mas apenas aos seus componentes.

Em suma, este artigo apresenta a proposta, implementação e avaliação de uma plataforma de inteligência analítica para o processo de participação social, a qual contempla todas as etapas tradicionais do processo de descoberta do conhecimento em bancos de dados. Em termos de contribuições específicas, podemos destacar:

- (1) Mecanismos para coletar, extrair e armazenar dados de participação social;
- (2) Projeto, implementação e validação de uma plataforma para carga e seleção de dados de participação social;
- (3) Projeto, implementação e validação de modelos e algoritmos de inteligência analítica aplicados a dados de participação social;
- (4) Projeto, implementação e validação de uma interface interativa e visual para inteligência analítica aplicada a dados de participação social.
- (5) Publicação da plataforma como software livre, incluindo modelos e algoritmos de inteligência analítica.
- (6) Utilização da plataforma para analisar as discussões em torno do projeto de Proteção a Dados Pessoais.

2 ARQUITETURA DA SOLUÇÃO

Nesta seção apresentamos a arquitetura da plataforma para acompanhamento de participação social através da internet, a qual é ilustrada na Figura 1 (Seção 2.1). Em seguida, na Seção 2.2, descrevemos o modelo de operação e funcionamento da plataforma, que é descrito através das tarefas que são executadas na plataforma. Mais informações podem ser obtidas no sistema Web disponível online⁶.

2.1 Arquitetura da Plataforma

O ponto de partida da arquitetura (ver Figura 1) é base de dados da consulta, obtida através de uma exportação do sítio que hospeda a mesma. Essa base de dados é então carregada em uma base de dados da plataforma (“Base de dados”), a qual suporta todas as atividades de inteligência analítica da plataforma. Essa base de dados deve suportar toda sorte de consultas de seleção, provendo as respostas a estas consultas no formato JSON e gerando “Dados selecionados”. Os dados selecionados são a entrada das “Tarefas analíticas” a serem executadas sobre os dados.

A execução das tarefas pode gerar três tipos de informações: (i) indicar a necessidade de alteração nos parâmetros de seleção, com vistas a novo ciclo analítico; (ii) enriquecimento dos dados, os quais são incorporados à “Base de dados”; e (iii) metáforas visuais, as quais auxiliam o entendimento dos dados e do comportamento dos usuários no processo de participação social.

2.2 Modelo de Operação e Funcionamento

O Modelo de operação e funcionamento da plataforma desenvolvida é orientado a dados, pois assim fica flexível, expansível e escalável. Assim a arquitetura da plataforma, descrita na Seção 2.1, funciona

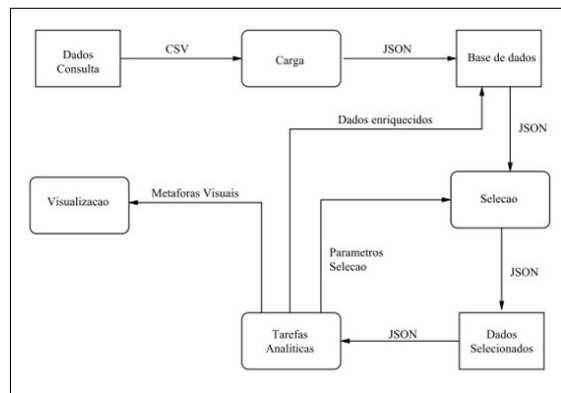


Figura 1: Arquitetura do Sistema

por meio de tarefas que são executadas de forma independente, seguindo o modelo apresentado na Figura 2.

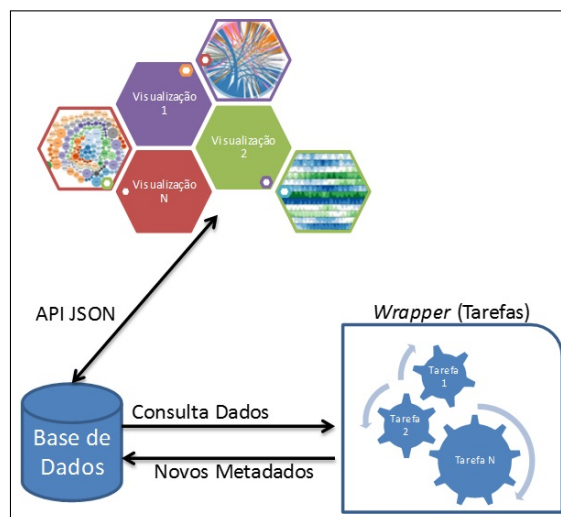


Figura 2: Arquitetura Operacional - Modelo de Dados

Os elementos básicos desta arquitetura são:

- (1) Base de dados: contém todos os dados da consulta pública, que são os seus itens comentáveis de um projeto de lei, os comentários (posts dos usuários) e os respectivos usuários participantes da consulta.
- (2) Wrapper de tarefas: consiste de um módulo que empacota as tarefas de enriquecimento de dados, contendo uma entrada de dados, parâmetros de execução e saída (ou resultado do processo) que agrega valor aos dados, adicionando a informação de uma das seguintes formas:
 - Um novo campo ou coluna como metadado aos dados existentes;
 - Um novo registro que adicione o metadado e relacione o mesmo a 2 ou mais entidades das existentes no modelo de dados (item comentável, usuário ou comentário).

⁴www.knime.com/

⁵orange.biolab.si/

⁶<http://staging-ceweb.speed.dcc.ufmg.br/>

- (3) Visualização de dados: módulo que engloba as técnicas que visualização de dados que consomem direto da base de dados através de uma API JSON.

3 ESTUDO DE CASO - ANTEPROJETO DE LEI PROTEÇÃO DE DADOS PESSOAIS

O Anteprojeto de Lei de Proteção de Dados Pessoais é resultado de um amplo debate público promovido pelo Ministério da Justiça, em parceria com o Observatório Brasileiro de Políticas Digitais do Comitê Gestor da Internet no Brasil, que teve duração de cinco meses, recebendo mais de 14 mil visitas e obteve mais de 800 contribuições. Esses subsídios foram analisados e discutidos no âmbito do Poder Executivo entre os órgãos interessados e, posteriormente, consolidados em um texto final. Este texto foi submetido novamente a debate público a partir do dia 28 de janeiro de 2015, sendo finalizado no dia 5 de julho de 2015. Qualquer indivíduo, após se registrar na plataforma www.participacao.mj.gov.br pode contribuir comentando os artigos do anteprojeto de lei em discussão ou enviando contribuições consolidadas em PDF.

A proposta visa assegurar ao cidadão o controle e a titularidade sobre suas informações pessoais, de modo a concretizar o direito à inviolabilidade da intimidade e da vida privada, protegido constitucionalmente. O avanço da tecnologia da informação amplia enormemente o potencial de coleta, processamento e utilização de dados pessoais, o que representa, por um lado, uma oportunidade de geração de novos conhecimentos e serviços, mas, por outro lado, pode acarretar graves riscos aos direitos da personalidade do cidadão, ao acesso a serviços e bens, além de uma grande insegurança jurídica para o ambiente de negócios de tecnologia da informação existente no país.

A utilização cada vez mais intensa de dados pessoais na sociedade da informação cria um desequilíbrio entre os poderes dos indivíduos, titulares de seus próprios dados pessoais, e os dos utilizadores de tais dados, justamente pela quantidade de informações pessoais que as novas tecnologias são capazes de agregar e utilizar. Para que esses dados possam ser utilizados com fins transparentes e legítimos, ao mesmo tempo em que sejam garantidos os direitos de seus titulares, são necessárias normas e mecanismos institucionais que estabeleçam os parâmetros e limites deste manejo, até mesmo no momento de término dessa relação. Além disso, tendo em vista o caráter transnacional do fluxo dessas informações, cumpre indicar que qualquer discussão que verse sobre este tema abrange tanto as operações de tratamento de dados pessoais realizadas no Brasil, como aquelas realizadas no exterior, mas em que a coleta tenha ocorrido em território nacional.

Dada a importância não só do tema, mas a relevância da construção de metodologias de análises de participação social na construção legislativa que se faz cada vez mais presente na construção de diversas políticas públicas, há a necessidade de se pensar como rotular, validar e incorporar todas as informações de forma objetiva e analítica.

A seguir apresentamos as principais funcionalidades da ferramenta COPPER, aplicada aos dados reais da consulta pública de Proteção de Dados Pessoais.

3.1 Funcionalidades e Análise

3.1.1 Matriz de correlação de comentários. Esta análise leva em conta o processamento de padrões textuais determinando um modelo que seja capaz de indicar um **índice de interesse**, isto é, buscase calcular a **densidade** dos comentários de um grupo de cidadãos em um conjunto de artigos de anteprojeto de lei.

Representação Visual

Os Artigos de lei estão dispostos em colunas e os grupos de cidadãos estão definidos em linhas. Cada visualização apresenta uma lista com diversos modelos de padrões de relacionamentos entre os comentários dos cidadãos e as diversas seções dos artigos de lei. Cada modelo de padrões gera uma matriz de coocorrência diferente.

A matriz de uma coocorrência dos comentários (ver Figura 3) de um conjunto cidadãos muda levando em conta o conjunto de artigos de lei comentados que a compõe de acordo com o interesse dos comentários dos cidadãos. Ex.: Um conjunto de artigos de leis podem ter um maior número de comentários por apenas um conjunto específico de cidadãos. Logo, a matriz de coocorrência desse subconjunto difere da matriz de coocorrência global, pois o índice de interesse identificado é diferente. A legenda apresenta a distribuição do número de comentários num intervalo de tonalidade de cores. O interesse dos cidadãos aos artigos está definido pela cor do quadrado, sendo a densidade de comentários proporcional a tonalidade.

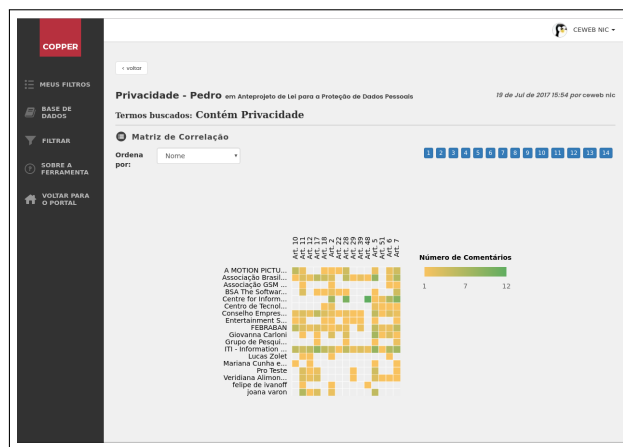


Figura 3: Matriz de Coocorrência de termos

3.1.2 Grafo de co-ocorrência dos conjuntos de termos frequentes. Esta análise apresenta a frequência da relação de ocorrência entre as palavras ou sequências lógicas filtradas pelo usuário com as demais palavras da base de dados sob a condição de que todas estejam interligadas.

Representação Visual Cada ocorrência é representada por nós e sua área é proporcional a frequência de ocorrência. As arestas retratam a coocorrência entre as palavras, e sua espessura é proporcional a frequência dessa coocorrência, conforme exemplo das Figuras 4 e 5.

Os grafos de co-ocorrência são separados em duas visualizações. **Grafo de Co-ocorrência:** A interligação de palavras co-ocorrentes

apenas entre o conjunto de dados que satisfazem as palavras ou relacionamentos lógicos do filtro realizado pelo usuário.

Grafo de Co-ocorrência Completo: A interligação de palavras co-ocorrentes entre o conjunto de dados que satisfazem as palavras ou relacionamentos lógicos do filtro realizado e todos os outros termos mais frequentes da base de forma que todos os elementos estejam interligados.

As palavras que foram utilizadas no filtro que satisfazem a relação de co-ocorrência são destacados em nós brancos. Ao clicar duas vezes sob qualquer nó, seus relacionamentos são destacados. A espessura da aresta representa a frequência dos correlacionamentos. Ao passar o mouse por cima do nó, a palavra relacionada fica em destaque apresentando o número elementos interligados, ou seja seus vizinhos.

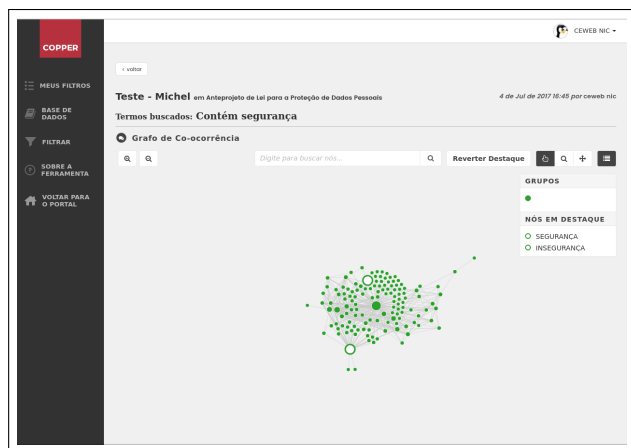


Figura 4: Co-ocorrência de termos (Grafo de Co-ocorrência) - Os elementos do filtro ficam em destaque.



Figura 5: Co-ocorrência de termos (Grafo de Co-ocorrência) - Quando um termo é selecionado seus relacionamentos ficam em destaque.

3.1.3 Grafo de Participação (Participantes e Artigos). Esta análise apresenta a interligação entre os participantes e os artigos comentados.

Representação Visual O grafo entre os participantes e os artigos comentados forma um dígrafo (ver Figura 6), isto é, não existem arestas que orientem o relacionamento entre participantes ou artigos isoladamente, um participante só se relaciona com artigos e vice-versa. Os participantes estão dispostos em losangos e os artigos em círculos verdes, a aresta entre eles representa o número de comentários. Quanto maior a espessura de uma aresta, maior o número de comentários realizados por um participante em um artigo. Ao passar o mouse por cima de qualquer nó, aparece o respectivo nome (artigo ou participante) juntamente com número de vizinhos, sendo os mesmos destacados da representação. Para tornar esse destaque persistente, basta clicar duas vezes sobre o elemento escolhido.

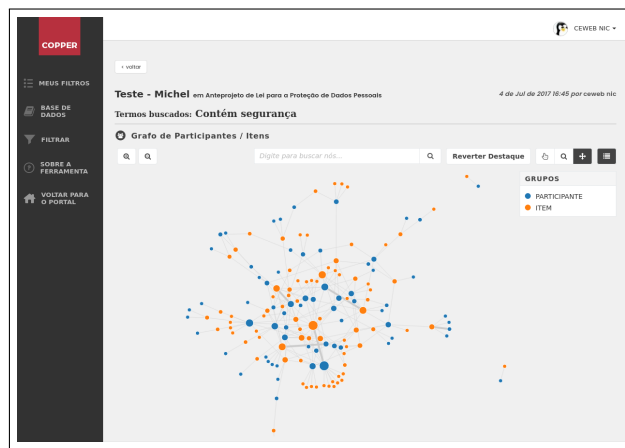


Figura 6: Grafo de participação

3.1.4 Visualização Árvore de Palavras. Esta análise apresenta uma representação em árvore (grafo) da frequência de palavras ou seqüências em um texto, conforme Figura 7. A árvore de palavras é formada por todas as palavras da base que satisfazem os relacionamentos lógicos utilizados no filtro de busca, para o caso da utilização do filtro.

Representação Visual O tamanho das palavras destacadas na árvore está vinculado sua a frequência na base. As frases, ou o conjunto de palavras relacionadas estão ligadas por arestas. Ao clicar na palavras esses nós podem ser restringidos, possibilitando ao usuário uma visualização mais específica.

3.1.5 Análise de Sentimento. Esta análise leva em conta o processamento de padrões textuais determinando um modelo que é capaz de indicar o quanto um comentário indica uma opinião positiva de um participante em um artigo de lei.

Representação Visual Os artigos estão dispostos em colunas e os participantes estão dispostos em linhas, conforme ilustra a Figura 8 A cor do quadrado define a probabilidade de uma opinião positiva dos comentários. A legenda apresenta o percentual dessa probabilidade distribuída no intervalo de cores. Quanto mais próximo a 1, maior a probabilidade de que os comentários representam uma opinião positiva do participante. A visualização classifica os participantes e os artigos de acordo com a positividade dos comentários.



Figura 7: Árvore de Palavras

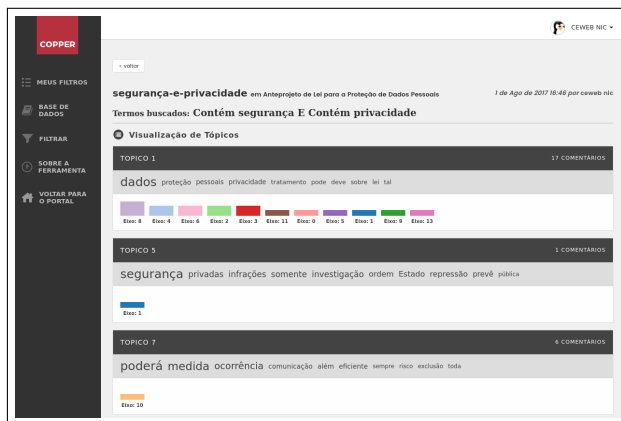


Figura 9: Visualização de Tópicos

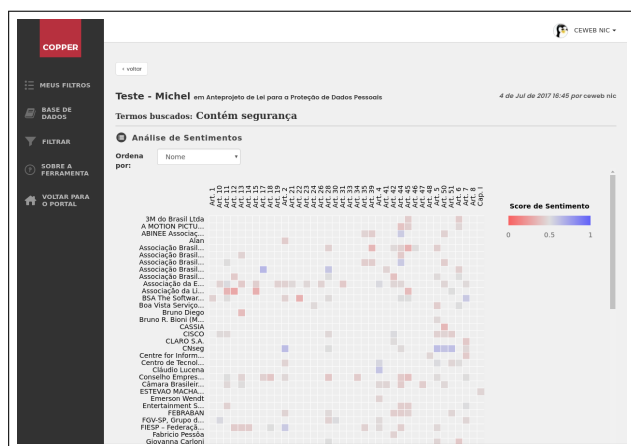


Figura 8: Matriz de Análise de Sentimento

3.1.6 *Visualização de Tópicos*. Esta análise tem por objetivo sumarizar os comentários referentes a um mesmo tópico de interesse, identificando o grupo de termos descritivos que representam seu conteúdo semântico. Os documentos dos quais os tópicos são extraídos, são classificados em eixos de debates.

Representação Visual Cada linha da tabela representa o tópico o qual os documentos pertencem, conforme Figura 9. O tamanho das barras representa o número de comentários recebidos por cada documento, sumarizados na coluna comentários. A coluna termos mais descritivos reúne as palavras que mais representam o conteúdo semântico do tópico de interesse. O tamanho dos termos apresenta a representatividade da palavra dentro do conteúdo semântico do tópico.

4 METAS DE APLICAÇÃO E LICENÇA

Dentre as metas de aplicação deste projeto, pretende-se aplicar as ferramentas a diferentes bases de dados de consultas públicas do nosso governo federal. Em especial, já foi aplicada à consulta

pública de proteção de dados pessoais e também bases de dados do portal e-democracia, pertencente ao Congresso Nacional.

Essa aplicação tem o código livre, baseado na licença de uso e distribuição *Creative Commons*.

5 CONCLUSÃO

Este trabalho apresentou a proposta, implementação e avaliação de uma plataforma de inteligência analítica para o processo de participação social ou consulta pública, contemplando todas as etapas tradicionais do processo de descoberta do conhecimento em bancos de dados.

Foram especificadas e implementadas diferentes funcionalidades inerentes ao processo de descoberta de conhecimento ou ciência de dados. Como validação real da ferramenta desenvolvida, utilizou-se a nova plataforma para analisar as propostas e discussões em torno do projeto de Proteção a Dados Pessoais. Tal aplicação possibilitou formalizar o método de uso e os benefícios de tal uso para a e-democracia.

Como trabalhos futuros, pretendemos utilizar outras bases reais de consultas públicas, bem como gerar novas técnicas de ciência de dados para enriquecer as análises de dados de tais mecanismos de e-democracia.

AGRADECIMENTOS

Este trabalho foi parcialmente patrocinado pelo NIC.br / CEWEB.br / W3C Brasil, INCT para a Web (CNPq no. 573871/2008-6), MASWeb (grant FAPEMIG/PRONEX APQ-01400-14), EUBra-BIGSEA (H2020-EU.2.1.1 690116, Brazil/MCTI/RNP GA-000650/04), CAPES, CNPq, Finep, e Fapemig.

REFERÊNCIAS

- [1] H. Catt and M. Murphy. What voice for the people? categorising methods of public consultation. *Australian Journal of Political Science*, 38(3):407–421, 2003.
- [2] K. Culver and P. Howe. Calling all citizens: The challenges of public consultation. *Canadian Public Administration*, 47(1):52–75, 2004.
- [3] P. Levy. *Ciberdemocracia*. Instituto Piaget, 2003.
- [4] F. P. J. Marques. *Ciberpolítica: Conceitos e Experiências*. Editora da Universidade Federal da Bahia, 2016.