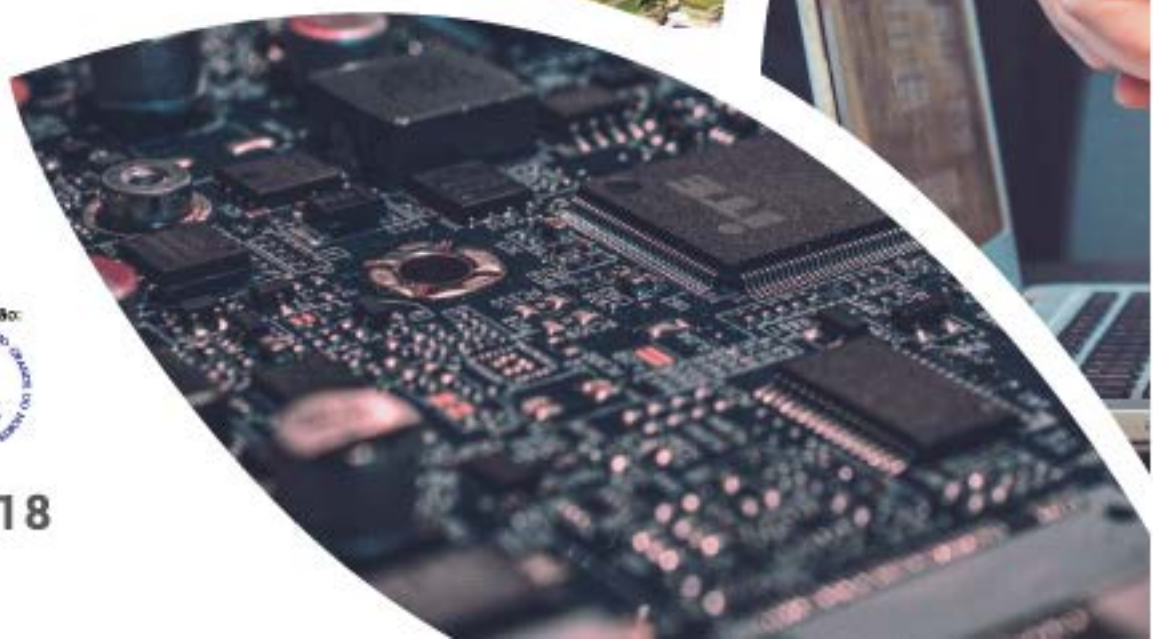


anais 2018

XXXVIII CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO
6º WTRANS – WORKSHOP DE TRANSPARÊNCIA EM SISTEMA
CENTRO DE CONVENÇÕES | NATAL•RN | 22 A 26 DE JULHO DE 2018
#COMPUTAÇÃOESUSTENTABILIDADE



NATAL, 2018

cnqis 2018

XXXVIII CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO
CENTRO DE CONVENÇÕES | NATAL•RN | 22 A 26 DE JULHO DE 2018
#COMPUTAÇÃOESUSTENTABILIDADE



Coordenador Geral

Francisco Dantas de Medeiros Neto (UERN)

Comissão Organizadora

Bartira Paraguaçu Falcão Dantas Rocha (UERN)

Camila Araújo Sena (UERN)

Everton Ranielly de Sousa Cavalcante (UFRN)

Felipe Torres Leite (UFERSA)

Ilana Albuquerque (UERN)

Isaac de Lima Oliveira Filho (UERN)

Priscila Nogueira Krüger (UERN)

Realização

Sociedade Brasileira de Computação

Organização

Universidade do Estado do Rio Grande do Norte

CSBC 2018

XXXVIII Congresso da

Sociedade Brasileira de Computação

Apresentação

Estes anais registram os trabalhos apresentados durante o XXXVIII Congresso da Sociedade Brasileira de Computação (CSBC 2018), realizado em Natal-RN, de 22 a 26 de julho 2018. O evento teve como tema central a Computação e Sustentabilidade, pois se compreende que o avanço da computação e as questões ambientais devem caminhar lado-a-lado, tendo em vista que as técnicas computacionais necessitam ser usadas para possibilitar o desenvolvimento sustentável, e, desse modo, equilibrar as necessidades ambientais, econômicas e sociais.

Organizar o maior evento acadêmico de Computação da América Latina foi um privilégio e um desafio. Foi enriquecedor promover e incentivar a troca de experiências entre estudantes, professores, profissionais, pesquisadores e entusiastas da área de Computação e Informática de todo o Brasil. Ao mesmo foi desafiador termos que lidar, principalmente, com às dificuldades impostas pelo momento de crise que o nosso Brasil vem enfrentando. Uma crise que afeta diretamente nossas pesquisas e, conseqüentemente, o desenvolvimento e inovação do nosso amado Brasil.

Por meio de seus 25 eventos, o CSBC 2018 apresentou mais de 300 trabalhos, várias palestras e mesas-redondas. O Congresso ainda abrigou diversas reuniões, que incluem a reunião do Fórum de Pós-Graduação, a reunião do CNPq/CAPES, a reunião dos Secretários Regionais SBC, a reunião das Comissões Especiais e a reunião do Fórum IFIP/SBC.

O sucesso do CSBC 2018 só foi possível devido à dedicação e entusiasmo de muitas pessoas. Gostaríamos de agradecer aos coordenadores dos 25 eventos e aos autores pelo envio de seus trabalhos. Além disso, gostaríamos de expressar nossa gratidão ao Comitê Organizador, por sua grande ajuda em dar forma ao evento; e, em especial, à equipe da Sociedade Brasileira de Computação (SBC), por todo apoio.

Por fim, reconhecemos a importância do apoio financeiro da CAPES, do CNPq, do CGI.br, do Governo do Estado do Rio Grande do Norte, da Prefeitura Municipal do Natal, da Prefeitura Municipal de Parnamirim, da CABO Telecom, da ESIG Software e Consultoria, da DynaVideo e do SENAI.

Natal (RN), 26 de julho de 2018.

Chico Dantas (UERN)
Coordenador Geral do CSBC 2018

Anais do CSBC 2018

**6º WTRANS – WORKSHOP DE
TRANSPARÊNCIA EM SISTEMA**

Coordenação Geral

Vanessa Nunes (UnB)
Bruna Diirr (UNIRIO)

Coordenação Local

Sebastião Emidio (UERN)

Comitê de Programa

Antônio de Pádua A. Oliveira (UERJ)
Ana Cristina Bicharra (UFF)
Bruna Diirr (UNIRIO)
Célia Ghedini Ralha (UnB)
Claudia Cappelli (UNIRIO)
Clayton Slaviero (UFMT)
Cristiano Maciel (UFMT)
Eduardo Almentero (UFRRJ)
Flávia Bernardini (UFF)
Gustavo Almeida (UFF)
Henrique Prado Sousa (UFRRJ)
José Viterbo (UFF)
Joanna Pivatelli (PUC-Rio)
Julio Cesar Sampaio do Prado Leite (PUC-Rio)
Luiz Marcio Cysneiros (York University/Canada)
Marilia Ferreira (PUC-Rio)
Milene Serrano (UnB-FGA)
Patrícia Vilain (UFSC)
Priscila Engiel (PUC-Rio)
Vera Maria Benjamim Werneck (UERJ)

SUMÁRIO

Formação para Cidadania e Controle Social via Educação a Distância	7
Cristiano Maciel, Cassyra Vuolo, Alexandre Anjos, Taciana Sambrano, Ana Paula Kuhn, Claudia Rouiller	
Análise de Sentimento no Contexto de Comentários a Projetos de Lei Relativos à Transparência	17
Roxana Portugal, Julio Leite	
Monitoramento automático da qualidade das informações que contribuem para a transparência pública	27
Fabricio Massariol, Fabiano da Rocha Louzada	
Integração e Análise das Despesas do Governo Federal	37
Diego Pasqualin, Luis Carlos De Bona, Marcos Sunye, Cristian Weiland, Edemir Maciel	
Uso de banco de dados orientado a grafos na detecção de fraudes nas cotas para exercício da atividade parlamentar	43
Gabriel Araujo, Maristela Holanda	
Transparência para Humanos e Máquinas: Um framework para Publicar Dados Abertos Interconectados Semanticamente Descritos	53
Julio Dos Reis, Mario Brito	
A iniciativa da construção do Plano de Dados Abertos do Tribunal de Contas do Rio Grande do Norte através de parceria acadêmica com a Universidade Federal do Rio Grande do Norte	63
Frederico Pranto Filho, Renan Silva, Lindemberg Silva Pereira, Joir Ramalho	
Desafios Sociotécnicos e Prospecções para Promover Transparência de Dados na Universidade	73
Breno de França, Julio Dos Reis, Rodolfo Azevedo	
Aplicando Transparência no Domínio de Informação Sobre Saúde na Internet: Uma Abordagem Visando o Cidadão	83
Nathalia Miranda, Henrique Sousa, Eduardo Almentero, Julio Leite	

Formação para Cidadania e Controle Social via Educação a Distância

Cassyra Lucia Correa Barros Vuolo¹, Cristiano Maciel², Alexandre Martins dos Anjos², Ana Paula Kuhn², Taciana Mirna Sambrano², Claudia Oneida Rouiller¹

¹ Tribunal de Contas do Estado de Mato Grosso
Cuiabá – MT – Brazil

² Universidade Federal de Mato Grosso (UFMT)
Cuiabá – MT – Brazil

{cassyra, claudia}@tce.mt.gov.br, {crismac, dinteralexandre}@gmail.com,
{profapaulaunemat, tacianamirna}@hotmail.com

Abstract. *Brazilian councils for the management of public policies are spaces for the exercise of citizenship, which requires preparation. This paper presents the results of the union between different public government institutions to offer an online course on Citizenship and Social Control. This initiative of education for transparency was intended to integrate the government and the civil society, as well as to foster transparency and enable society to perform the social control. The scientific publicization of the lessons learned in this initiative, which reached about 700 citizens by means of a Learning Management System, can spur new actions and reflections in this area.*

Resumo. *O exercício da cidadania requer preparo, o qual, na estrutura brasileira, tem um espaço privilegiado nos conselhos de gestão das políticas públicas. Neste sentido, este artigo apresenta os resultados da união entre diferentes instituições governamentais públicas para realização de um curso de “Cidadania e Controle Social” via Educação a Distância. Tal iniciativa de educação para transparência serve para aproximação e integração do governo e da sociedade civil, bem como de estímulo a transparência e capacitação da sociedade para realização do controle social. A divulgação científica das lições apreendidas nesta iniciativa, que atingiu cerca de 700 cidadãos, via ambiente virtual de aprendizagem, pode fomentar novas ações e reflexões nesta área.*

1. Introdução

Para que possamos ter uma democracia participativa, é necessário a participação dos cidadãos na vida pública (MACIEL, GARCIA, 2007). Segundo Nerling (2017), a comunidade, a cidadania, a participação, a transparência e o controle permeiam as diversas etapas do ciclo de gestão de políticas públicas. Todavia, o exercício da cidadania requer preparo, o qual, tem um espaço privilegiado na estrutura brasileira dos conselhos de gestão das políticas públicas (NERLING, 2017).

No que se refere ao controle social, é uma forma de compartilhamento do poder de decisão entre o Estado e sociedade em relação às políticas. É um instrumento da democracia e da cidadania. E pode ser realizado tanto no momento da implantação de políticas públicas quanto no momento de fiscalização (KUHN et al., 2017). Para Kuhn

et al. (2017), o controle interno, por sua vez, tem por finalidade o acompanhamento das funções das atividades dos órgãos subordinados à administração, observando os princípios constitucionais estabelecidos no âmbito da gestão pública. Por certo que os conselhos não são os únicos instrumentos de democratização e controle social sobre a gestão dos recursos públicos, mas são os que possuem na sua composição, de forma paritária, representantes do governo e da sociedade e têm, por natureza, o dever de propor e avaliar os resultados das políticas públicas.

É nessa conjuntura que ocorre a articulação entre o Tribunal de Contas de Mato Grosso (TCE-MT) com os conselhos de políticas públicas, foco desta pesquisa. O TCE-MT, enquanto o órgão responsável pela fiscalização dos resultados da gestão pública, com capilaridade sobre os dados relacionados às políticas públicas para elevar a efetividade e a transparência de suas decisões, pode empoderar os conselhos para a realização de controle social (GIRATA et al., 2013; OLIVEIRA; MACIEL, 2013). Somente com informações e conhecimento de suas funções na democracia participativa serão possíveis o acompanhamento e a fiscalização do poder público pelos conselhos de políticas públicas.

Face ao exposto, o presente artigo apresenta os resultados da união de esforços entre diferentes instituições governamentais públicas por meio da adoção da Educação a Distância como ferramenta pedagógica de aproximação, integração e capacitação da sociedade para realização do controle social. Esta experiência exitosa teve início em 2012 com esforços conjuntos de órgãos governamentais. Na ocasião, o Tribunal de Contas do Estado de Mato Grosso (TCE-MT) tomou a decisão de firmar parceria com a Universidade Federal do Estado de Mato Grosso (UFMT), por intermédio da Fundação de Apoio e Desenvolvimento da UFMT (Fundação Uniselva), para realizar o curso de extensão intitulado “Cidadania e Controle Social” via Educação a Distância (EaD), reconhecendo a universidade pública como lócus de conhecimento tecnológico e formação para a cidadania política (VUOLO; MACIEL, dos ANJOS, 2017).

Segundo Maciel et al. (2016), sugere-se na “área de Educação, entre outros, o desenvolvimento de soluções para desenvolver a educação para a transparência e o incentivo ao uso de dados abertos governamentais na forma gráfica como meio de atrair a atenção e facilitar o entendimento dos dados pelos cidadãos de modo a elevar e melhorar o nível de participação democrática no país.”. Os autores sugerem um conjunto de ações a serem realizadas, também para a área de “Educação Digital”. A educação para transparência é algo que se faz necessário e premente. Na literatura, em especial no Brasil, são escassos os trabalhos nesta área, todavia percebem-se iniciativas neste contexto educativo como a dos autores Monsalve, Werneck e Leite (2013), em um estudo de transparência especificamente com jogos digitais na área de Pedagogia.

No curso resultante da parceria supracitada, o objetivo foi permitir a inclusão cidadã nos processos participativos de gestão a partir da disponibilização de informações e de conhecimentos que favorecessem o empoderamento social e a interface entre os controles externo e social. O curso foi realizado no Ambiente Virtual de Aprendizagem (ANJOS, 2014) Moodle. A análise do processo e dos resultados da iniciativa mediada pelas tecnologias contribui para o debate sobre a relevância da educação corporativa permitindo ampliar as reflexões sobre como esta pode ser oferecida para toda a sociedade.

A proposta do curso trouxe também para análise e compartilhamento a

contribuição possível e efetiva da união de esforços entre a academia e o controle externo. Teve como proposta final a elaboração de um trabalho factual relacionando às necessidades de sua comunidade e às políticas públicas propondo soluções possíveis, inovadoras e viáveis, geradoras de valor público coletivo. Foram apresentados aproximadamente 700 trabalhos e selecionados os cinquenta melhores textos, os quais foram objeto de análise no curso (VUOLO; MACIEL, dos ANJOS, 2017). Neste caso, as ponderações com pessoas de vários municípios e de diferentes realidades, viabilizando um diálogo democrático e a transferência de experiências, de conhecimento e de informações, só foram possíveis com a utilização do ambiente virtual de aprendizagem e a incorporação das tecnologias digitais.

Ao final, é possível entender o poder da união de diferentes conhecimentos e informações públicas que, mediados pela tecnologia do ambiente virtual de aprendizagem permitem, sob o ponto de vista educacional, o acesso e a atuação de diferentes sujeitos e a construção de novos conceitos e ideias que impregnam esses ambientes (VALENTE, 2002).

2. Cenário Político, Social e Legal

Para melhor compreensão dos fatos, será apresentado o cenário político, social e legal desta conjuntura de fatos e ações. Pode-se dizer que o ponto de partida para a realização dessas ações ocorreu a partir da promulgação da Constituição Federal de 1988 por meio da qual foram institucionalizados uma série de espaços para a participação da sociedade civil na gestão pública nos três níveis governamentais (municipal, estadual e federal).

Peruzzotti e Smulovitz (2002) que estudam a participação cidadã associada à vigilância dos recursos públicos, defendem que há uma nova forma de politização que está se desenvolvendo nas novas democracias e que tem por objetivo primordial fortalecer e aperfeiçoar o funcionamento dos mecanismos de controle e supervisão das instituições representativas, e que os cidadãos afetados pelas ações públicas, no exercício de sua cidadania, devem pedir contas aos gestores sobre suas ações.

A interlocução entre o controle externo e o controle social é recepcionada na consolidação dos direitos civis, sociais e políticos que convergem para a enumeração das oportunidades e dos instrumentos de democratização, transparência e inserção da voz do cidadão nos processos decisórios da gestão pública, principalmente no viés da orçamentação das políticas públicas.

A interação crescente entre os controles estatal (controle externo) e não estatal (controle social) foi sendo inserida na legislação brasileira, como na Lei de Responsabilidade Fiscal, a Lei Complementar n.º 101/00 (LRF), que determinou a necessidade de equilíbrio das finanças públicas, introduziu a garantia da transparência e da participação da sociedade no processo de elaboração das políticas públicas (BRASIL, 2000). A necessidade dessa aproximação ganhou reforço com a edição da Lei de Acesso à Informação – Lei 12.527/11, que tornou a transparência uma ação obrigatória para toda a administração pública bem como a implementação nos sítios oficiais eletrônicos, de espaços contendo todas as informações públicas de interesse coletivo ou geral produzidas ou custodiadas pelo Estado (BRASIL, 2011). Esta arquitetura legal implicou em vários desafios para as instituições públicas e a sociedade. Para as primeiras há o dever de implementar mecanismos formais de participação e promover o

empoderamento social das suas informações e decisões e, para a sociedade, a incumbência de exercer seu papel de principal interessada nos resultados do Estado, apta a demandar dele novas estruturas de governança que possibilitem o desempenho de funções de proposição, avaliação, direcionamento e monitoramento das políticas públicas. Segundo Torres (2008, p.180), a comunidade poderá invocar a proteção ao Tribunal de Contas para o combate à corrupção, para o controle dos incentivos fiscais e para promover a fiscalização sobre as entidades financeiras privadas.

De acordo com o Relatório de Desenvolvimento Humano (RDH) elaborado pelas Nações Unidas, em 2015, o Brasil figura entre as sociedades mais desiguais do mundo, aparecendo como o 10º no ranking, e suas políticas públicas precisam ser formuladas e executadas para amenizar esta realidade e gerar melhoria na qualidade de vida da população. É exatamente em função desse cenário que ocorreu a interface entre os cidadãos, o órgão de controle externo - o TCE-MT e a universidade – a UFMT.

3. Controle social e controle externo: a integração e os Tribunais de Contas

Em se tratando de controle, é válido ressaltar, a fim de se fundamentar a discussão que se pretende, que esta ação deve acontecer desde o planejamento, execução e fiscalização constante da destinação do dinheiro público. Acerca das prerrogativas constitucionais relacionadas ao controle dos recursos públicos, é importante ressaltar aqui o teor da Emenda Constitucional nº 19/98, que alterou o parágrafo 3º do Art. 37 da Constituição Federal. Ao tratar das formas de participação do cidadão/usuário na administração pública, a Emenda Constitucional nº 19/98 estabeleceu uma perspectiva de controle direto dos recursos públicos nos seguintes termos:

§ 3º– A lei disciplinará as formas de participação do usuário na administração pública direta e indireta, regulando especialmente: I – as reclamações relativas à prestação dos serviços públicos em geral, asseguradas a manutenção de serviços de atendimento ao usuário e a avaliação periódica, externa e interna, da qualidade dos serviços; II – o acesso dos usuários a registros administrativos e a informações sobre atos de governo, observado o disposto no art. 5º, X e XXXIII; III – a disciplina da representação contra o exercício negligente ou abusivo de cargo, emprego ou função na administração pública.

Assim, a Carta Magna brasileira possibilita ao cidadão exercer uma forma de democracia direta (de soberania popular). A materialização desse exercício caracteriza-se por ser o controle social. Os meios de controle social teriam como base de sustentação a fiscalização das ações públicas. Portanto, o controle social poderia ocorrer por monitoramento legal ou iniciativa individual ou coletiva de grupo social.

De acordo com Faccioni (2006), o Tribunal de Contas da União tem por primazia a promoção de uma ligação entre os indivíduos e a administração, na medida em que contribui para a efetivação da democracia participativa, o exercício da cidadania (enquanto princípio fundamental) e a interface entre o controle externo e o controle social. Nessa esteira, evidencia-se o Art. 74, § 2º da Constituição Federal de 1988, que atribuiu ao Tribunal de Contas da União a competência para apreciação de irregularidades ou ilegalidades na Administração Pública, a partir da ação do cidadão e algumas entidades, a saber, qualquer cidadão, partido político, associação ou sindicato é parte legítima para, na forma da lei, denunciar irregularidades ou ilegalidades perante o Tribunal de Contas da União.

Os Tribunais de Contas, enquanto fiscalizadores da gestão, devem mensurar os resultados obtidos pelo governo e a qualidade das políticas públicas, buscar que sua atuação tenha ressonância na sociedade e disponibilizar informações que precisam ser entendidas e consumidas pelo cidadão. Do lado das instituições, pode haver o receio de se relacionar com a sociedade, como assevera Fernandes (2008); da perspectiva da sociedade, pode haver a “letargia social” que precisa ser abandonada para “alavancar a expressão de uma consciência que se revela imanente no corpo social, aguardando motivação”(p.50).

4. O Curso de Extensão: “Cidadania e Controle Social”

O curso foi ofertado inicialmente em 2013 com apenas 200 vagas. Em 2017 foram ofertadas 2.000 vagas do curso de EAD – “Cidadania e Controle Social”, divididas em 40 turmas, acompanhadas por 40 tutores e supervisionadas por quatro técnicos do TCE-MT, certificado pela UFMT como curso de extensão. O presente estudo abordará apenas a versão do curso desenvolvida em 2017.

O objetivo fundamental do curso foi capacitar os conselheiros dos Conselhos Municipais e Estaduais, estudantes e membros da sociedade para o exercício do controle social, fomentando sua interface com o controle interno e externo. Os objetivos específicos do curso foram a. compreender controle social como pressuposto da democracia; b. contribuir para o fortalecimento dos Conselhos Municipais, Estaduais e Escolares como instâncias efetivas de controle social, levando-os a propor formas de intervenção em sua realidade local; c. conhecer e caracterizar as três formas de controle da Gestão Pública (interno, externo e social), identificando os atores responsáveis; d. Conhecer o TCE-MT, os projetos e programas desenvolvidos relacionados ao estímulo ao controle social como o PDI, os sistemas de informação e transparência e as ações de controle social; e. Conhecer o site do TCE-MT e as decisões do controle externo relacionadas às políticas públicas de sua cidade; f. Conhecer o ciclo orçamentário e as formas de participação do cidadão no processo de orçamentação pública.

Contando com um aparato de fórum e uma estratificação em coordenadores, tutores e alunos, o acompanhamento passou a ser feito pontualmente através de reuniões, relatórios e outras formas. O curso contemplou 100 horas dedicadas à discussão sobre a importância e as diferentes formas de acompanhamento e controle social e externo, com vistas à definição e vivência da cidadania. A aprovação ficou condicionada à elaboração de um trabalho conclusivo substancial pelos cursistas. As produções acadêmicas oriundas do curso representam o indicador de maior representatividade de importância do curso ofertado.

4.1. Sobre o curso

Desta vez, mais de 1.000 alunos concluíram o curso, o que representou uma taxa de evasão de 50%, sendo que os 679 alunos foram aprovados e redigiram um trabalho conclusivo substancial, o que lhes viabilizou a certificação pela UFMT (MEC). Esta produção acadêmica veio a ser o indicador de maior representatividade da nova fase do curso de extensão.

Entre as inovações executadas na versão do curso por ora apresentado, algumas foram muito positivas como: a) a abertura à população em geral, desde que se possuísse pelo menos ensino fundamental completo; por isso, as turmas eram compostas por

cursistas de perfis e conhecimentos heterogêneos, residentes na capital, no interior de Mato Grosso e até em outros estados da Federação; b) o novo layout do AVA e o design gráfico dos módulos com ícones de fácil identificação prenderam a atenção do cursista no menu “Texto Didático”, bem como o “Guia de Estudo”, o que favoreceu um melhor entendimento das regras gerais de funcionamento de toda estrutura do curso; o acompanhamento das interações realizadas no AVA possibilitaram avaliar a quantidade de acesso e a participação crítica/reflexiva no ambiente; c) a oportunidade para os cursistas assistirem ao vivo, via Facebook, ao evento Consciência Cidadã e interagirem simultaneamente, comentando ou fazendo perguntas aos participantes desse evento de estímulo ao controle social foi bem avaliada; d) a inserção dos menus “Saiba Mais”, com bibliográficas suplementares, e do “Vídeos”, com temas correlatos aos debatidos no curso, propiciou a compreensão do assunto através de outra linguagem, bem como estimulou a ampliação dos horizontes dos cursistas; e) novas funcionalidades inseridas no ambiente, entre as quais destacou-se a inserção de alguns “menus” para auxiliar os cursistas no desenvolvimento do curso, o “Suporte Técnico”, os “Tutoriais” e o “FAQ”, destinados a sanar dúvidas dos cursistas por meio da compilação das perguntas mais recorrentes e/ou dúvidas possíveis acerca do curso; f) a “Sala do Tutor” integrou os coordenadores, supervisor, tutores, suporte técnico da UFMT/Uniselva e coordenação do TCE, e no “Fórum Geral do Aluno” as dúvidas dos cursistas sobre os assuntos teóricos e exercícios do curso foram tiradas com mais celeridade.

Segundo Mill (2010), nos últimos anos, a Educação a Distância brasileira passou por uma vertiginosa expansão e por um enorme reconhecimento nunca percebido na modalidade. Nesse contexto de expansão da EaD, insere-se o curso apresentado no presente trabalho. Os números mostram que o TCE-MT, em parceria com a UFMT, atendeu 97% dos 141 municípios do Estado de Mato Grosso. Assim, tivemos cursistas de 137 municípios participando efetivamente do Curso. Além desses, foram atendidos 26 municípios de outros estados. Salienta-se que o maior número de cursistas inscritos concentrou-se nos municípios mais populosos, como Cuiabá e Várzea Grande. Entre os inscritos, verificam-se cursistas com nível fundamental de escolaridade, bem como doutores que desempenham suas atividades como defensores públicos, fiscais de tributos estaduais, serviços gerais, do lar, entre outros.

Ao final do Curso, os alunos apresentaram o trabalho final com reflexões importantes sobre o percurso realizado durante os estudos dos respectivos módulos. Ao todo, tivemos aproximadamente 700 (setecentos) trabalhos. Desses, 50 foram previamente selecionados, observando-se a estrutura quanto à formatação do trabalho, problematização, coerência e coesão textual (VUOLO; MACIEL, dos ANJOS, 2017). Neste quadro, pode-se identificar que os 50 trabalhos escolhidos são provenientes de 27 municípios do estado de Mato Grosso, sendo que, destes, 38% são da capital mato-grossense (19 trabalhos). A análise realizada sobre os artigos selecionados evidencia que as temáticas abordadas foram extremamente variadas. Parece haver uma preocupação maior com as políticas públicas voltadas à preservação ambiental e à educação, esporte e lazer. Salientamos que os trabalhos analisados envolvendo políticas públicas educacionais enfatizavam alunos da educação básica. A preocupação com a atuação junto aos Conselhos Municipais também aparece de forma significativa.

No Figura 1, há as temáticas abordadas para a construção dos trabalhos finais do Curso de Cidadania e Controle Social. São pesquisas voltadas para as Políticas Públicas dos municípios nos quais estavam inseridos os cinquenta trabalhos selecionados.

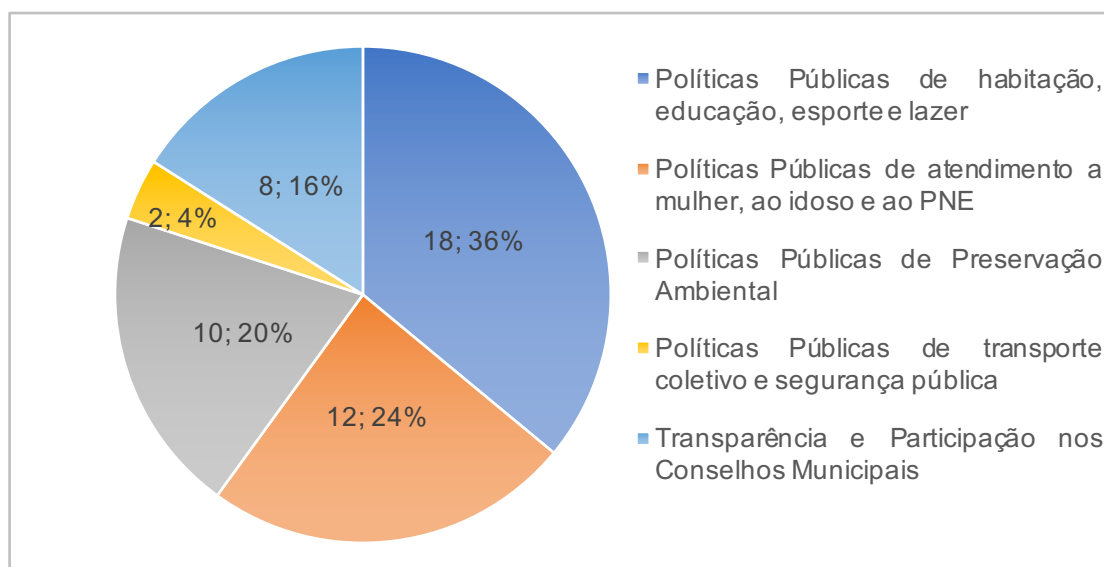


Figura 1. Temáticas abordadas nos trabalhos dos participantes do curso

Na representação gráfica, é possível perceber a preocupação com diversas temáticas quando se fala em Políticas Públicas. A maior preocupação trazida nos textos reflexivos refere-se às políticas de habitação, educação, esporte e lazer. Viu-se que 35% das pesquisas selecionadas demonstram a necessidade da garantia das condições mínimas de sobrevivência quando se fala em cidadania. Talvez seja importante enfatizar que ter essa consciência perpassa todo um processo de sensibilização, de empoderamento das questões sociais em busca da garantia dos direitos e consequentemente do exercício da cidadania plena.

Outra preocupação bastante evidente nos textos refere-se às políticas públicas de atendimento à mulher, ao idoso e ao portador de necessidades especiais. Identifica-se 25% dos trabalhos envolvidos com essa temática. As discussões realizadas ao longo do curso possibilitaram uma sensibilização com as minorias. É importante salientar que a temática saúde apareceu significativamente nas reflexões dos trabalhos voltados à mulher e ao idoso, bem como reflexões pertinentes sobre a acessibilidade dos que possuem alguma necessidade especial. Os trabalhos mostram como o atendimento à saúde tem sido precário especialmente quando se reporta à mulher e ao idoso.

Nos 20% dos trabalhos que tratam de políticas públicas voltadas à preservação ambiental, percebe-se uma grande preocupação com a destinação dos resíduos sólidos. Os trabalhos apontam que ainda não há efetivamente políticas públicas eficazes que consigam resolver o problema da destinação dos resíduos sólidos e de saneamento básico. Ainda nesse contexto das temáticas abordadas, temos 16% dos trabalhos preocupados com a Transparência e fiscalização dos conselhos Municipais. Os dados coletados apontam que a população ainda não tem consciência da importância da participação efetiva nos diversos conselhos existentes.

Percebe-se que, nesse contexto, a transparência ganha expressão como um tema estratégico da agenda pública do país e se consagra como instrumento fundamental para se trabalhar com a complexa relação entre os cidadãos e seus agentes. A utilização efetiva de ferramentas e tecnologias de informação, especialmente da internet, torna-se importante aliada e apresenta-se como processo democrático que integra uma rede

capaz de dar visibilidade e transparência às decisões e aos resultados alcançados, através da criação de condições favoráveis de acesso aos dados públicos. Os outros 4% dos trabalhos preocuparam-se com o transporte e a segurança pública. A preocupação maior esteve voltada para a qualidade do transporte público e a segurança nos pontos de ônibus. Além da necessidade do fortalecimento de políticas para proteção do usuário do transporte público, as pesquisas mostram a necessidade de priorizar os deslocamentos coletivos nos espaços urbanos para diminuir o tempo de viagem da população.

Ainda, cabe reforçar que os números totalizados no projeto são expressivos: 2.567 pessoas foram estimuladas a exercerem o controle social por meio desta capacitação, havendo representantes em 100% dos municípios adesos ao PDI, além de 22 cursistas de outros estados da federação. Além disso, dos 676 cursistas que concluíram o curso, ou seja, que obtiveram nota igual a 7,0 ou superior, 41% são membros de Conselhos, 54% deles ligados aos Conselhos da Educação, tanto que houve 113 professores concluintes. Dos 141 municípios de Mato Grosso, 122 possuíam municípios inscritos e em 89 deles houve concluintes.

5. Considerações finais

As inovações, de forma geral, contribuíram em muito para o sucesso desta versão remodelada desta iniciativa de educação para transparência, sendo que o trabalho da Escola de Contas merece ser destacado na medida em que não envidou esforços para realizar com a habitual excelência este curso que não se encaixava nos padrões dos cursos ofertados pelo ensino à distância.

Nesse sentido, o curso de Cidadania e Controle Social trouxe importantes reflexões tais quais: o que é ser cidadão? Quem sou eu na sociedade? De que forma posso contribuir para melhorar as condições sociais do meu país e exercer efetivamente a cidadania? Como realizar uma denúncia e acompanhar as ações tomadas na resolução dos problemas? Entende-se que a conscientização é o ponto de partida para que o cidadão possa repensar e transformar sua práxis. Nesse sentido, salienta-se a importância do Curso de Cidadania e Controle Social na preparação dos conselheiros que fazem parte dos conselhos e do cidadão comum que atua, interage e pode intervir na sociedade de forma a garantir seus direitos e estar consciente de seus deveres. As práticas realizadas durante o desenvolvimento do Curso contribuíram significativamente para que essas reflexões e tomadas de decisões fossem pensadas de forma consciente e transparente, pautadas nas legislações existentes e visando sempre o bem comum.

Espera-se que, bem informados, os membros da sociedade deixem de ser administrados e tornem-se cidadãos ativos e influam de fato nos atos do poder tomados em seu nome. A integração ativa e participativa dos cidadãos, influenciando e decidindo na administração, é um dos pressupostos para a incorporação da sociedade civil como um dos intérpretes da Constituição (MANGIA, 2014).

O histórico de inovação e superação construído por todos os atores envolvidos na construção deste projeto educacional e profundamente social reflete, em sua íntegra, a mesma linha inovativa e progressiva do próprio Tribunal de Contas de Mato Grosso que, ao longo de seu amplo desenvolvimento institucional alcançado ao longo da última década, soube criar também uma vasta e profícua rede de articulação social. No caso específico deste curso de extensão, foi a rede criada entre TCE-MT, Secretaria de Articulação Institucional e Desenvolvimento da Cidadania – SAI, Escola Superior de

Contas, UFMT/Uniselva e os conselhos de políticas públicas que permitiu tamanho índice de protagonismo, de onde surge a certeza de que o caminho para uma democracia real e participativa passa, via de regra, pelo exercício do Controle Social.

Os dados informam que no processo de amadurecimento da democracia participativa, tanto a sociedade quanto o Estado são chamados a se modificarem, a construir aliança em torno de pautas coletivas, a praticarem mediações sociais e políticas visando o atendimento de pautas populares e a caminharem na direção da estruturação de esferas públicas autônomas e democráticas no campo das decisões políticas (RAICHELIS, 2000).

Sob o ponto de vista de empoderamento social, a diversidade de formação dos cursistas nos permite dizer que diferentes espaços de discussão plural foram atingidos de diferentes modos, sendo, inclusive, recomendado um estudo focal para identificar tais impactos e medir a efetividade da voz perante a gestão pública. Em um país com dimensões como as do Brasil, com cultura de má gestão dos recursos, ainda que a politização da sociedade esteja aumentando, somente a atuação dos órgãos de controle é insuficiente para combater as irregularidades cometidas, sendo cada vez mais urgente ouvir a sociedade, estimulá-la ao exercício do controle social e incentivar a educação para a transparência. Esperamos que esta experiência fortaleça o diálogo democrático e republicano, ainda pouco efetivado, mas possível e necessário, que pretende se lançar daqui para novos horizontes

Referências

- ANJOS, A. M. dos. Tecnologias da informação e da comunicação, aprendizado eletrônico e ambientes virtuais de aprendizagem. In Maciel, C. Educação a distância: Ambientes virtuais de aprendizagem. Cuiabá: EdUFMT, 2014, p. 11-57.
- BRASIL. XXIII ASSEMBLEIA GERAL DA OLACEFS. Revista do Tribunal de Contas da União. Brasil, ano 46, n. 129, p. 17, jan/abr., 2014. Disponível em <<http://revista.tcu.gov.br/ojs/index.php/RTC/article/download/18/16>>. Acesso em: 05 jun.2017.
- BRASIL. Lei nº 12.527, de 18 de novembro de 2011. Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei no 8.112, de 11 de dezembro de 1990; revoga a Lei no 11.111, de 5 de maio de 2005, e dispositivos da Lei no 8.159, de 8 de janeiro de 1991; e dá outras providências. http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm. Acesso em 10 jan 2018.
- BRASIL. Emenda Constitucional nº 19, de 04 de junho de 1998. Modifica o regime e dispõe sobre princípios e normas da Administração Pública, servidores e agentes políticos, controle de despesas e finanças públicas e custeio de atividades a cargo do Distrito Federal, e dá outras providências. Disponível em http://www.planalto.gov.br/ccivil_03/Constituicao/Emendas/Emc/emc19.htm. Acesso em 10 jan 2018.
- BRASIL. Constituição da República Federativa do Brasil de 1988. Disponível em http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em 10 jan 2018.
- BRASIL. Lei Complementar nº 101, de 4 de maio de 2000. Estabelece normas de finanças públicas voltadas para a responsabilidade na gestão fiscal e dá outras

- providências. Disponível em: http://www.planalto.gov.br/ccivil_03/Leis/LCP/Lcp101.htm. Acesso em 10 jan de 2018.
- DE OLIVEIRA, L. K. B.; MACIEL, C.. Transparência das Contas Públicas e Controle Social via Portal do Cidadão-um estudo de caso. Anais do WTRANS13 - Workshop de Transparência em Sistemas, Brasília, Brasil, Setembro 29, 2013.
- FACCIONI, V. J. A Sociedade e o Controle Externo. In: I Fórum do PROMOEX: Formação da Rede de Informação e Portal do Controle Externo. TCE-MG. Belo Horizonte, 2006.
- GIRATA, N. N. H.; MACIEL, C.. eGov website evolution study within strategic planning. In: Proceedings of the 15th Annual International Conference on Digital Government Research. ACM, 2014. p. 69-78.
- KUHN, A. P.; ROUILLER, C. O.; MACIEL, C.. Um olhar sobre os trabalhos dos participantes do curso a distância de cidadania e controle social. In VUOLO, Cassyra L.; MACIEL, Cristiano; ANJOS, Alexandre dos. Educação a Distância: (In)formação em Cidadania e Controle Social. Cuiabá-MT: EdUFMT, 2017, p.173-188.
- MACIEL, C., GARCIA, A. C. B.. Modeling of a Democratic Citizenship Community to Facilitate the Consultative and Deliberative Process in the Web. In ICEIS, 2007, pp. 387-394.
- MACIEL, C., CAPPELLI, C., SLAVIERO, C., GARCIA, A. C. B.. Technologies for popular participation: a research agenda. In: Proceedings of the 17th International Digital Government Research Conference on Digital Government Research. ACM, 2016. p. 202-211.
- MANGIA, C. de C.. Participação popular, cidadania e Estado Democrático de Direito. Revista Jus Navigandi. Teresina, ano 19, n. 4094, 16 set. 2014. Disponível em: <<https://jus.com.br/artigos/29663>>. Acesso em: 26 dez. 2016.
- MILL, D.. Das inovações tecnológicas às inovações pedagógicas: considerações sobre o uso das tecnologias na educação a distância. In: MILL, D. PIMENTEL, N. Educação a Distância: desafios contemporâneos. São Carlos:EdUFSCar, 2010.
- MONSALVE, E.; WERNECK, V.; LEITE, J. C. S. P. Incorporando transparência na pedagogia através do uso de jogos para ensino. Anais do WTRANS13 - Workshop de Transparência em Sistemas, Brasília, Brasil, Setembro 29, 2013.
- NERLING, M.. Preparar para o exercício da cidadania, através do controle social, na Gestão de Políticas Públicas. IN: VUOLO, C. L; MACIEL, C.; ANJOS, A. M. dos. Educação a Distância: (In)formação em Cidadania e Controle Social. Cuiabá-MT: EdUFMT, 2017, p. 17-67
- PERUZZOTTI, E.; SMULOVITZ, C. Controlando la Política: ciudadanos y médio sem las nuevas democracias. Buenos Aires: Editorial Temas, 2002, p. 23-52.
- RAICHELIS, R.. Desafios da gestão democrática das políticas sociais. Módulo 03. Capacitação em Serviço Social e Política Social. Programa de Capacitação Continuada para Assistentes Sociais. Brasília, CFESS, ABEPSS, CEAD/NED-UNB, 2000.
- TORRES, R. L.. Curso de Direito Financeiro e Tributário. 7. ed. Rio de Janeiro: Renovar, 2008.
- VALENTE, J. A. A espiral da aprendizagem e as tecnologias da informação e comunicação: repensando conceitos. In: Joly, M. C. (Ed) Tecnologia no Ensino: implicações para a aprendizagem. São Paulo: casa do psicólogo, 2002, p. 15-37.
- VUOLO, C. L; MACIEL, C.; ANJOS, A. M. dos. Educação a Distância: (In)formação em Cidadania e Controle Social. Cuiabá-MT: EdUFMT, 2017.

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	18%

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: *Obrigar 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 Casas 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 trabalho 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. Azarado, 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

Monitoramento automático da qualidade das informações que contribuem para a transparência pública

Fabricio Massariol, Fabiano da Rocha Louzada

Secretaria de Estado de Controle e Transparência do ES
Av. Governador Bley, 236, Centro – Vitória/ES – Brasil

{fabricio.massariol, Fabiano.louzada}@secont.es.gov.br

***Abstract.** With the application of the Law on Access to Public Information, in 2012, transparency portals have been expanded to comply with this legislation. However, not all public managers have sufficient resources to analyze the information that is being uploaded daily. Providing information without quality will not meet the basic principle of transparency, that of making society understand information and participating in government actions. This article demonstrates how to operationalize a data analysis of a transparency portal and how this operation to improve the quality characteristics of the information to the public.*

***Resumo.** Com a vigência da Lei de Acesso à Informação, em 2012, os portais de transparência têm sido ampliados para atendimento a esta legislação. Contudo, nem todos os gestores públicos possuem recursos suficientes para analisar as informações que estão sendo publicadas diariamente. Disponibilizar a informação sem qualidade não atenderá ao princípio básico da transparência, o de fazer com que a sociedade entenda as informações e participe dos atos governamentais. Este artigo demonstra como operacionalizar a análise dos dados de um portal da transparência e como esta operação contribui para melhorar as características de qualidade das informações disponíveis ao público.*

1. Introdução

É notório o avanço no uso de novas tecnologias para disponibilização de informações à sociedade, porém, a credibilidade nas informações disponibilizadas pelos portais de dados abertos e de transparência ainda é um grande desafio, principalmente em países em desenvolvimento, conforme destacado por Dabla-Norris *et al.* (2010) e Gollwitzer (2010).

Parte desse problema é devido à heterogeneidade das informações publicadas, sua forma de atualização e frequência. A Modernização e Ampliação do Portal da Transparência do Governo do Estado do Espírito Santo [Massariol, Louzada and Alves, 2017] demonstra que para atendimento à Lei de Responsabilidade Fiscal [Brasil, 2000] e Lei de Acesso à Informação [Brasil, 2011], foi necessário buscar dados em diversos sistemas da administração pública, alguns provenientes de bancos de dados, integrados de forma automática e atualizados diariamente, outros, produzidos em planilhas e atualizados manualmente pelos respectivos responsáveis.

Manter dados e informações disponíveis aos cidadãos, provenientes de vários sistemas, com frequência de atualização e integrações específicas, de forma íntegra, correta

e atualizada, exige um esforço grande de validações por parte do gestor público que, muitas vezes, não possui recursos suficientes para realizar tal procedimento.

Este artigo irá demonstrar como operacionalizar a análise dos dados de um portal da transparência, por um *web crawler*, para atendimento às características de qualidade da informação. Para tanto, será usado o robô MAILA (acrônimo de Monitoramento Automático de Informações de Livre Acesso), utilizado no Portal da Transparência do Governo do Estado do Espírito Santo, e os trabalhos desenvolvidos por Parker *et al.* (2006) e Cappelli (2009), que fornecem um conjunto de características que contribuem para a transparência das informações.

Segundo Reis (2013), um *web crawler* é um programa de computador que busca e recupera informações da *Web* de forma automática, “visitando” suas páginas e armazenando as informações em um repositório central para que sejam analisadas para diversos propósitos, por exemplo, para suportar um mecanismo de busca geral, um mecanismo de busca especializado em um determinado assunto, para análises de inteligência de negócios (coletar informações sobre empresas concorrentes, parceiros de negócio, notícias relevantes), para o monitoramento de *web sites* de interesse, etc.

Na seção 2 é apresentada a fundamentação teórica; na seção 3, as características e funcionamento do robô MAILA; na seção 4, as características da qualidade da informação verificadas; na seção 5, seus benefícios e resultados obtidos até o momento; na seção 6, conclusão e perspectivas futuras.

2. Fundamentação Teórica

O primeiro sítio eletrônico designado de Portal da Transparência foi elaborado em 2004, pela Controladoria Geral da União (CGU) (2018), com a finalidade de publicar informações orçamentárias e financeiras, conforme estabelece a Lei de Responsabilidade Fiscal – LRF [Brasil, 2000]. O Estado do Espírito Santo elaborou e publicou seu portal em 2009, sob a responsabilidade da Secretaria de Estado de Controle e Transparência, de acordo com os prazos estabelecidos pela Lei Complementar nº 131/2009 [Brasil, 2009], que determina a disponibilização, em tempo real, de informações pormenorizadas sobre a execução orçamentária e financeira da União, dos Estados, do Distrito Federal e dos Municípios.

No ano de 2011 foi publicada a Lei de Acesso à Informação – LAI, nº 12.527/2011 [Brasil, 2011], que estabeleceu a diretriz que todos os documentos são públicos e, o sigilo, a exceção. Neste momento, o rol taxativo de informações descritas pela LRF, passou a ser o rol mínimo exigido pela LAI, além da exigência em ter um portal amigável e que pudesse alcançar o maior número possível de cidadãos. Os desafios impostos por esta nova legislação, levou o Governo do Estado do Espírito Santo a elaborar um projeto de modernização e ampliação do Portal da Transparência, com informações e funcionalidades bem mapeadas e que pudessem abranger os três perfis de públicos-alvo mapeados pela pesquisa [Massariol, Louzada and Alves, 2017].

O resultado do trabalho foi a ampliação do número de consultas disponíveis, provenientes de vários sistemas do Poder Executivo e do Tribunal de Contas, com diversas formas de integração e frequências de atualização, conforme listado no Quadro 1 a seguir:

Assunto	Responsável	Sigla do	Origem dos dados /	Frequên
---------	-------------	----------	--------------------	---------

		Sistema	Ferramenta de integração	cia
Despesa e receitas	Secretaria da Fazenda	SIGEFES	Arquivos CSV / <i>Hang fire</i>	Diária
Licitações e Contratos	Secretaria de Recursos Humanos	SIGA	Banco SQL-SERVER / <i>Integration Server</i>	Diária
Obras	Tribunal de Contas	Geo-Obras	Banco Oracle / <i>Hang Fire</i>	Diária
Leis orçamentárias	Secretaria de Planejamento	-	Sítio institucional / <i>Web Service</i>	Anual
Convênios e Transferências	Secretaria da Fazenda	SIGEFES	Arquivos CSV / <i>Hang Fire</i>	Diário e Mensal
Cartão corporativo	Secretaria de Recursos Humanos	-	Arquivos CSV / <i>Hang fire</i>	Mensal
Bolsa Capixaba	Secretaria de Assistência Social	Sistema da CEF	Arquivos CSV / <i>Hang fire</i>	Mensal
Patrimônio	Secretaria de Recursos Humanos	SIGA	Banco SQL-SERVER / <i>Integration Server</i>	Diária
Servidores Públicos	Secretaria de Recursos Humanos	SIARHES	Banco Oracle / <i>Hang Fire</i>	Diária
Prestação de Contas do Governador	Secretaria de Controle e Transparência e Tribunal de Contas	-	Arquivos em PDF	Anual
Saúde	Secretaria de Saúde	Sistema de indicadores	Arquivo XML / <i>Web Service</i>	Diária
Segurança	Secretaria de Segurança Pública	SIAT	Banco Oracle / <i>Hang Fire</i>	Diária
Educação	Secretaria de Educação	SIGES	WebService / <i>Hang Fire</i>	Semanal
Conselhos	Secretaria de Governo	-	Planilha excel / atualizado pelo Usuário	mensal
Mapas	Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural	GeoBases	<i>View</i> no Banco Oracle / <i>Hang Fire</i>	Diário

Quadro 1 - Consultas disponíveis no Portal da Transparência do ES

No estudo em que Parker *et al.* (2006) desenvolveram, foi constatado que um modelo de qualidade da informação para aplicação em portal de transparência na *web*, deve apresentar algumas das seguintes dimensões, a seguir: acessibilidade, atualidade, exatidão, relevância, credibilidade, objetividade, completeza, representação, fonte, compreensibilidade.

Além disso Parker *et al.* (2006) fornecem direções claras e mensuráveis, resumindo as dimensões apresentadas por outros autores para avaliação da qualidade das informações.

A Acessibilidade (*Accessibility*) refere-se ao quão facilmente disponíveis estão as informações no site e para download; a Acurácia (*Accuracy*) refere-se a quão confiáveis e corretas estão as informações na página *web*; a Adequação (*Appropriateness*) refere-se a quão adequado é o conteúdo, de acordo, com as necessidades dos visitantes da páginas *web*; a Credibilidade (*Believability*) ao quão verdadeiro e credível é o conteúdo na página; a Integridade (*Completeness*) em relação a falta de dados no conteúdo da *web* e se são suficiente; A Consistência (*Consistency*) ao fato de as páginas *web* serem apresentadas no mesmo formato; a Facilidade de Manipulação (*Ease of Manipulation*), ou seja, o quão fácil de manipular é o conteúdo da página; Livre de erro (*Free-of-Error*) refere-se ao fato de as informações sobre a página *web* serem corretas e confiáveis, ou seja, livre de erros; a Objetividade (*Objectivity*) refere-se ao quão imparcial e sem preconceitos é o conteúdo da página *Web*; a Relevância (*Relevancy*) refere-se a análise do quão aplicável, útil e relevante é o conteúdo da página *Web*; a Representação (*Representation*) do quão legível, consistente e estruturado formalmente é o conteúdo da página *Web*; a Reputação (*Reputation*) da informação na página da *web* é altamente considerada no que diz respeito ao seu conteúdo. A Segurança (*Security*), ou seja, em que medida o acesso à página da *Web* é restrito de forma adequada para manter a sua segurança; a Velocidade (*Speed*) de recuperação ou para download do conteúdo de páginas *web*; a Pontualidade (*Timeliness*) com a qual o conteúdo da página *Web* é atualizado; a Compreensibilidade (*Understandability*) se o conteúdo da página *Web* é facilmente compreendido; e, o Valor acrescentado (*Value-added*), ou seja, se as informações sobre a página são benéficas e fornecerem vantagens na sua utilização [Parker *et al.*, 2006].

Em um trabalho correlato, Cappelli (2009) propõe a utilização de um *Softgoal Interdependency Graph* (SIG) [Chung *et al.*, 2000] para representar o conceito de transparência organizacional, contendo algumas as características desejáveis. Carvalho, Cappelli e Santoro (2017), por sua vez, utilizam o SIG de transparência proposto por Cappelli (2009) para sugerir uma abordagem de avaliação de transparência em sites. As características do SIG de transparência de Cappelli (2009) são relacionadas a seguir:

- Acessibilidade, que é subdividida em portabilidade (capaz de ser utilizado em diversos ambientes), disponibilidade (capaz de aceitar solicitações), e publicidade (capaz de ser conhecido);
- Usabilidade, que é subdividida em uniformidade (capaz de manter uma única forma), intuitividade (capaz de ser utilizado sem aprendizado prévio), simplicidade (capaz de não apresentar dificuldades ou obstáculos), amigabilidade (uso fácil), operabilidade (capaz de ser executado), e desempenho (capaz de operar adequadamente).
- Informativo, que é subdividida em clareza (capaz de prover informações com nitidez), acurácia (ausência de erros), completeza (capacidade de não faltar nada do que pode ou deve ter), corretude (capacidade de ser isento de erros), consistência (capacidade de resultar aproximadamente em várias medições de um mesmo item), integridade (capacidade de ser preciso e rigoroso), comparabilidade (capacidade de ser confrontado com outro para lhe determinar diferença, semelhança ou relação), e atualidade (refletir a última atualização ou mudança).

- Entendimento, que é subdividida em compositividade (capacidade de construir e formar a partir de diferentes partes), concisão (capacidade de ser resumido), divisibilidade (capacidade de ser particionado), e adaptabilidade (capacidade de mudar de acordo com as circunstâncias e necessidade).
- Auditabilidade, que é subdividida em explicável (capaz de explicar dados, procedimentos e regras), rastreabilidade (capaz de seguir o desenvolvimento de um processo ou a construção de uma informação, suas mudanças e justificativas), verificabilidade (capaz de ser legitimado, identificar se o que está sendo feito é o que deve ser feito), validade (capaz de ser testado por experimentação ou observação, identificar se o que está sendo feito é correto), e controlabilidade (capaz de ser controlado).

3. Monitoramento Automático de Informações de Livre Acesso – MAILA

Devido à ampliação de informações que vem ocorrendo no Portal da Transparência do Estado do Espírito Santo, e da necessidade de mantê-lo atualizado aos cidadãos, surgiu a necessidade de se criar um robô para auxiliar o processo de análise dos dados, contribuindo para melhorar a qualidade das informações disponíveis aos cidadãos.

“No que toca a organização ou gestão da informação, a quantidade de dados e informações disponíveis tem levado as organizações a implementarem ações para a otimizar a coleta, o tratamento, a localização, análise das informações e avaliação” [Queyras and Quoniam, 2006, p.81].

O desenvolvimento do robô MAILA iniciou no ano de 2014, especificamente para controlar os dados de despesa, pois, à época, ocorriam diversos problemas tanto na extração quanto na atualização dessas informações no portal da transparência. Com o passar dos anos, devido à ampliação das informações disponíveis ao público, houve a necessidade também de se ampliar as validações.

O robô MAILA foi construído em *PowerShell*, que é uma interface de linha de comando para acessar os serviços dos sistemas operacionais da Microsoft, onde podem ser elaborados *scripts* sobre a plataforma .NET [Júnior and Gomes, 2013]. O robô, atualmente, está dividido em 3 seções específicas: leitura dos dados; análise dos dados; comunicação do resultado.

1. Leitura dos dados:

Com a reestruturação do portal da transparência, as consultas foram desenvolvidas para exibir o valor consolidado e a data em que ocorreu a última atualização dos dados, em formato HTML, permitindo que o script realize a leitura dessas informações. Exemplo: Na consulta de licitações, em seu cabeçalho, são exibidas a quantidade de licitações e o valor gasto no ano; na consulta de despesa, aparecem os valores empenhado, liquidado e pago realizados no ano; na consulta do programa social bolsa capixaba, aparecem a quantidade e o valor pago aos beneficiários no ano etc.

2. Análise dos dados:

Na etapa anterior, os valores consolidados disponíveis em cada consulta, juntamente com a data em que ocorreu a última atualização, são armazenados em uma planilha para serem analisados. Com a leitura desses dados, é possível verificar se as informações estão disponíveis para acesso ao público, analisar a tendência de crescimento ou redução com base no histórico e identificar possíveis anomalias, causadas por erros de implementação dos sistemas ou por falhas dos usuários. A análise da tendência de crescimento ou redução, e a identificação das anomalias, é feita mediante comparação dos dados históricos que estão armazenados na planilha. Por exemplo, se determinada informação apresenta histórico de crescimento diário de 5% e, em determinado dia, cresce 10%, trata-se de um indício de anomalia. O Quadro 2, a seguir, demonstra os problemas identificados pelo robô MAILA, como são apresentados aos usuários e possíveis causas:

Tipo do problema	Como é apresentado ao usuário	Possíveis causas
Informações indisponíveis para acesso	- as informações não são exibidas	Sistema de origem está em manutenção e se encontra <i>off-line</i>
		Sistema de origem sofreu mudança na estrutura de dados
		Falha na ferramenta de integração que transfere os dados.
		Falha na infraestrutura que garante a conexão entre os ambientes
		Falha na indexação dos dados (<i>Elastic Search</i>)
Informações desatualizadas	- as informações são apresentadas de forma desatualizadas	Todos as causas listadas no item acima
		Falha do usuário em não atualizar as informações na data prevista (esquecimento, férias, mudança de equipe, ocupado com outras atividades etc)
Baixo desempenho na consulta por informações	- as informações não são exibidas devido ao tempo de espera exceder o <i>timeout</i> do servidor <i>web</i>	Problema de infraestrutura
		Problema de banco de dados
		Problema no código fonte
Informações incorretas (anomalias)	- as informações são apresentadas com valores incorretos	Importação dos dados incompleta
		Indexação dos dados incompleta
		Duplicação dos dados
		Falha de controle nos sistemas de origem (ex.: campo numérico que aceita valores sem nenhuma restrição)

Quadro 2 - Problemas identificados pelo robô MAILA, como são apresentados aos usuários e possíveis causas

3. comunicação do resultado:

O resultado é comunicado aos gestores do Portal da Transparência pelo e-mail corporativo (transparencia@secont.es.gov.br) e registrado em planilha para análise e acompanhamento futuros. Os resultados comunicados também são utilizados pela equipe de gestão para ajudar na identificação de problemas no Portal e nos sistemas fonte dos dados.

4. Melhoria da qualidade da informação

Conforme mencionado anteriormente, o robô MAILA é uma ferramenta que contribui para analisar dados do portal da transparência em curto espaço de tempo, tarefa que levaria algumas horas se fosse realizada manualmente.

Atualmente, o Portal da Transparência possui 44 (quarenta e quatro) consultas principais, as quais exibem informações em formato HTML, arquivos para download no formato PDF/XLSX e links para outras páginas. São 30 (trinta) as consultas que são verificáveis, ou seja, que não são links para outras páginas, e, destas, 27 (vinte e sete) são monitoradas pelo robô MAILA, o que equivale a uma abrangência de 90% das consultas verificáveis no Portal.

Esse procedimento automatizado permite que os gestores tomem conhecimento e consigam agir de forma rápida, muitas vezes, evitando que o problema chegue aos usuários finais.

Ressaltamos que os usuários do Portal da Transparência, cidadãos, estudantes e outros profissionais, dificilmente teriam o discernimento de identificar se o valor apresentado, na cifra de milhões ou bilhões de reais, estão ou não corretos. Informações apresentadas incorretamente, podem estar sendo utilizadas por centenas de pessoas diariamente.

Por isso, o gestor público deve agir rapidamente e garantir a qualidade da informação disponibilizada na internet. No Quadro 03 fazemos uma relação entre os tipos de problemas identificados pelo robô MAILA, com as características de qualidade relatadas por Parker *et al.* (2006) e com as características do SIG de transparência de Cappelli (2009), que estão momentaneamente ausentes no Portal da Transparência. Cabe ressaltar que nem todas as características presentes no SIG de transparência de Cappelli (2009), que são utilizadas por Carvalho, Cappelli e Santoro (2017) na abordagem de avaliação de transparência em sites, estão relacionadas com a qualidade da informação presente no Portal da Transparência.

Além disso, apesar de algumas características do SIG de transparência em sites estarem presentes no Portal da Transparência, por exemplo, a característica “explicável”, de Capelli (2009), uma vez que o Portal possui manual, *help online* e perguntas frequentes, tais características não estão sendo abordadas neste trabalho, uma vez que não se alteram com frequência e são difíceis de serem analisadas pelo robô MAILA.

Tipo de problema identificado pelo robô MAILA	Características (Parker <i>et al.</i> , 2006)	Características (Cappelli, 2009)
Informações estão indisponíveis	Acessibilidade	Disponibilidade

para acesso		
Informações desatualizadas	Acurácia; Credibilidade; Pontualidade	Atualidade
Baixo desempenho na consulta por informações	Velocidade	Desempenho
Informações incorretas (anomalias)	Acurácia; Credibilidade; Livre de erro	Consistência; Corretude; Integridade

Quadro 3 - Relação entre os tipos de problemas identificados pelo robô MAILA com as características de qualidade de Parker et al. (2006) e Cappelli (2009)

5. Benefícios e Resultados

Automatizar o processo de análise dos dados do portal da transparência resulta em diversos benefícios, destacamos os principais:

- redução do tempo de análise dos dados;
- maior exatidão para analisar dados e encontrar anomalias;
- baixo investimento para realizar manutenções no código fonte;
- monitoramento contínuo, mesmo na ausência dos responsáveis;
- a execução do script pode ser feita diversas vezes ao dia;
- maior confiabilidade das informações para os usuários;

Como resultado principal, podemos citar a identificação de anomalias nos dados, ocasionados devido à falta de controle dos sistemas que processam e armazenam as informações.

Os exemplos a seguir foram identificados pelo robô MAILA por meio de uma análise histórica dos dados, pois, nestes casos, houve um aumento do valor acima da média esperada:

- Em julho de 2017, uma licitação na modalidade de pregão eletrônico, referente à compra de materiais de limpeza, foi publicada no Portal da Transparência com o valor de quase R\$ 9 bilhões (nove bilhões) de reais. O gestor responsável foi alertado e a inspeção detectou uma falha no sistema de compras governamentais.
- Em agosto de 2017, os valores de despesa, referentes ao empenho, liquidação e pagamento, foram duplicados no Portal da Transparência. A equipe técnica foi acionada e a inspeção detectou uma falha no carregamento dos dados de despesa.
- Em outubro de 2017, um contrato no valor de R\$ 62 bilhões (sessenta e dois bilhões) de reais foi apresentado no Portal da Transparência. O gestor responsável foi alertado e a inspeção detectou que o usuário informou o CNPJ da empresa no campo valor do contrato. A informação foi corrigida no sistema de compras governamentais.

Outros casos identificados pelo robô estão relacionados à desatualização das informações, e suas principais causas são: mudança de ambiente tecnológico, ausência do responsável, mudança de equipe no setor e manutenção do sistema de origem dos dados.

6. Conclusão

A transparência é um bom antídoto contra a corrupção, pois incentiva os gestores públicos a serem mais responsáveis em sua atuação e permite que a sociedade, de posse das informações, controle a ação dos governantes e fiscalize a aplicação do dinheiro público.

Manter a qualidade das informações disponíveis aos cidadãos é essencial para alcançarmos a finalidade da transparência pública, por isso, a ampliação de informações nos portais deve ser acompanhada de um monitoramento contínuo dos gestores públicos, para tanto, se faz necessário a utilização de ferramentas para auxiliar a análise dos dados, com o objetivo de realizar as devidas correções em curto espaço de tempo.

A utilização do robô MAILA, para monitorar as informações que são publicadas no Portal da Transparência, tem permitido que o Governo do Estado do Espírito Santo disponibilize informações públicas com qualidade e atualizadas para a sociedade com menor gasto de tempo e de recursos.

Entretanto, algumas limitações foram identificadas, como a operacionalização do robô para analisar dados que estão em outros sítios institucionais ou para verificar a qualidade de informações que não se alteram com frequência e são difíceis de serem identificadas. Como exemplo, podemos citar a característica “explicável”, de Capelli (2009), uma vez que o Portal possui manual, *help online* e perguntas frequentes, que não se alteram com frequência e, quando são alteradas, são difíceis de serem analisadas pelo robô MAILA, pois são informações que abordam conceitos que dependem de análise de um especialista.

Finalizando, externamos as perspectivas futuras de evoluir o robô MAILA, incluindo novas funcionalidades, tais como:

- Confrontar os dados do diário oficial com o portal da transparência, para identificar informações que não estão nos sistemas de governo e, conseqüentemente, não estão disponíveis aos cidadãos pelo portal da transparência.
- Aumentar a granularidade na verificação dos dados, identificando se o gasto de um elemento de despesa ocorreu variação relevante, contribuindo, assim, para futuras inspeções e auditorias.

Referências

Brasil. (2000) “Lei Complementar nº 101, de 4 de maio de 2000”. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/lcp/lcp101.htm. Acesso em: 28/03/2018.

Brasil. (2009) “Lei Complementar nº 131, de 27 de maio de 2009”. Disponível em: http://www.planalto.gov.br/CCivil_03/leis/LCP/Lcp131.htm. Acesso em: 28/03/2018.

Brasil. (2011) “Lei nº 12.527, de 18 de novembro de 2011”. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/12527.htm. Acesso em: 28/03/2017.

Carvalho, L. P., Cappelli, C., and Santoro, F. M. (2017) “Transparência de Software Centrada na Análise de Sites.” In Anais do WTRANS17 - Workshop de Transparência em Sistemas, São Paulo, Brasil, Julho 3, 2017.

Cappelli, C. (2009) “Uma Abordagem para Transparência em Processos Organizacionais Utilizando Aspectos”. Tese de Doutorado. Pontifícia Universidade Católica do Rio de Janeiro. Rio de Janeiro.

CGU – Controladoria Geral da União. (2018). Transparência Pública. Disponível em: <http://www.transparencia.gov.br/sobre>. Acesso em 26 de março de 2018.

Chung, L., Nixon, B. A., Yu, E., Mylopoulos, J. (2000) “Non-Functional Requirements in Software Engineering.” Kluwer Academic Publishers, Londres

Dabla-Norris, Era et al. (2010). *Budget institutions and fiscal performance in low-income countries*. IMF working papers, 1-56

Gollwitzer, S. et al. (2010). *Budget Institutions and Fiscal Performance in Low-Income Countries*. International Monetary Fund.

Júnior, L. C. B. and Gomes, J. C. P. (2013) “Sistema Repositório Wiki de Scripts”. Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, Rio de Janeiro.

Massariol, F.; Louzada, F. R.; Alves, L. F. M. (2017) “Modernização e ampliação do portal da transparência do Espírito Santo: Pensado pelo cidadão para o cidadão”. In: X Congresso CONSAD de Gestão Pública, 2017, Brasília/DF.

Parker, M. B., Moleshe, V., De la Harpe, R., and Wills, G. B. (2006). “An evaluation of Information quality frameworks for the World Wide Web”.

Queyras, Joachim; Quoniam, Luc. Inteligência competitiva (IC). In: TARAPANOFF, K.(Org.). Inteligência, informação e conhecimento em corporações. Brasília: IBICT, UNESCO, p. 73-97, 2006

Reis, T. (2013). Algoritmo rastreador Web especialista nuclear. Tese de Doutorado. Universidade de São Paulo. São Paulo.

Integração e Análise das Despesas do Governo Federal

Cristian Weiland¹, Diego Pasqualin¹,
Edemir Maciel¹, Luis C. E. de Bona¹, Marcos Sunye¹

¹Centro de Computação Científica e Software Livre
Departamento de Informática
Universidade Federal do Paraná (UFPR)
R. Cel. Francisco H. dos Santos, 100 – Curitiba – PR – Brasil

{cw14, dpasqualin, bona, sunye, edemir}@inf.ufpr.br

Abstract. *Since 2012, with the creation of the law on access of information in Brazil, the amount of data available grew in such a way that the difficulty shifted from the access itself to the treatment and interpretation of the available information. In this context, SIMTransparência was developed with the objective of providing an alternative view of the Portal of Transparency of the Federal Government, displaying data from several datasets in an intuitive and dynamic interface, capable of assisting the public manager and brazilian citizens to better understand and control expenditures of the federal government. This article describes the technologies within SIMTransparência and ways to replicate it's datasets.*

Resumo. *Desde 2012, com o início da vigência da Lei de Acesso a Informação, a quantidade de dados disponíveis cresceu de tal forma que a dificuldade passou a ser não o acesso em si, mas o tratamento e interpretação das informações disponibilizadas. Nesse contexto o SIMTransparência foi desenvolvido, com o objetivo de proporcionar uma visualização alternativa ao Portal da Transparência do Governo Federal, exibindo dados de várias bases em uma interface intuitiva e dinâmica, capaz de auxiliar o gestor e cidadão comum a melhor compreender e fiscalizar as despesas do governo federal. Esse artigo descreve as tecnologias envolvidas no desenvolvimento do SIMTransparência e formas de replicar sua base de dados.*

1. Introdução

A Lei de Acesso a Informação (12.527/2011) garante a qualquer cidadão o acesso a informações sobre as atividades de órgãos públicos, sem a necessidade de identificação ou motivo. Desde o início de vigência da lei, em 2012, a quantidade de informação disponibilizada vem crescendo consideravelmente, reduzindo o problema de acesso aos dados, mas intensificando o problema de análise, a qual podemos considerar inviável sem o auxílio de ferramentas computacionais capazes de filtrar, manipular e destacar dados de interesse.

O SIMTransparência¹ é um projeto de Software Livre criado com esse propósito. Através da manipulação de uma grande quantidade de dados e exibição intuitiva em in-

¹Todo o código fonte, assim como documentação detalhada, encontra-se disponível sob licença GPL no repositório <https://gitlab.c3sl.ufpr.br/c3sl/transparencia>.

terface gráfica disponível via internet, ele busca aproximar cidadãos e gestores do processo de fiscalização dos gastos públicos. Atualmente, todos os dados disponibilizados pelo SIMTransparência são periodicamente coletados, tratados e filtrados a partir do Portal da Transparência do Governo Federal². Ao agrupar várias bases de dados e permitir aplicação de filtros avançados em diferentes escalas de tempo, o SIMTransparência se apresenta como uma ferramenta dinâmica, complementar ao Portal da Transparência e capaz de evidenciar despesas controversas e prioridades do órgãos do governo federal nos gastos públicos. Os dados coletados são tratados por *scripts* em *Shell* e *Python*, armazenados no banco de dados ElasticSearch³ e, por fim, visualizados em um navegador com a utilização da ferramenta Kibana³. Uma instância do projeto está disponível publicamente no endereço <http://www.c3sl.ufpr.br/transparencia>. Ela pode ser facilmente replicada e modificada em outros ambientes, processo este objetivo primário desse artigo.

As próximas seções descrevem o processo necessário para replicar o SIMTransparência em um ambiente GNU/Linux. A seção 2 detalha as bases de dados que serão utilizadas. As características do banco de dados são descritas na seção 3, enquanto que a seção 4 descreve o método sugerido para visualização dos dados armazenados. Em seguida, na seção 5, um resumo é apresentado, com os passos necessários para colocar o ambiente em produção, seguido por fim os desafios e limitações do sistema (seção 7) e a conclusão do trabalho.

2. Bases de Dados

Ao todo, sete conjuntos de dados do Portal da Transparência são utilizados: **i)** Diárias de Viagens⁴, concedidas ao servidor civil por dia de afastamento da sede do serviço, destinando-se a indenizá-lo por despesas extraordinárias com pousada, alimentação e locomoção urbana. **ii)** Cadastro de Servidores⁵, dados cadastrais de cada servidor, incluindo identificação por nome e CPF mascarado, cargo, função e atividades, entidade de lotação e de exercício, além do tipo e situação do vínculo, informações sobre afastamento (se houver), datas e documentos referentes ao ingresso e à jornada de trabalho. **iii)** Remuneração de Servidores⁵, corresponde remuneração e subsídio recebidos por ocupante de cargo e emprego público, incluindo auxílios, ajudas de custo, jetons e quaisquer outras vantagens pecuniárias, bem como proventos de aposentadoria e pensões. **iv)** Gastos Diretos⁶, despesas de custeio de manutenção das atividades dos órgãos da administração pública, como por exemplo: despesas com pessoal, juros da dívida, aquisição de bens de consumo, serviços de terceiros, manutenção de equipamentos, obras, água, energia, telefone, dentre outros. **v)** Lista de favorecidos por Natureza Jurídica⁷, lista contendo possíveis regimes jurídicos atribuídos a uma empresa ou agente autônomos (ex: autar-

²Site do Portal da Transparência: <http://portaltransparencia.gov.br/>.

³Site oficial da Elastic, empresa responsável pelo ElasticStack: <https://www.elastic.co/>.

⁴Diárias de viagens: <http://www.portaltransparencia.gov.br/downloads/mensal.asp?c=Diarias>.

⁵Cadastro de servidores e remuneração: <http://www.portaltransparencia.gov.br/downloads/servidores.asp>.

⁶Gastos diretos: <http://www.portaltransparencia.gov.br/downloads/mensal.asp?c=GastosDiretos>.

⁷Natureza jurídica, favorecidos e CNAE: <http://www.portaltransparencia.gov.br/downloads/mensal.asp?c=FavorecidosTransferencias>.

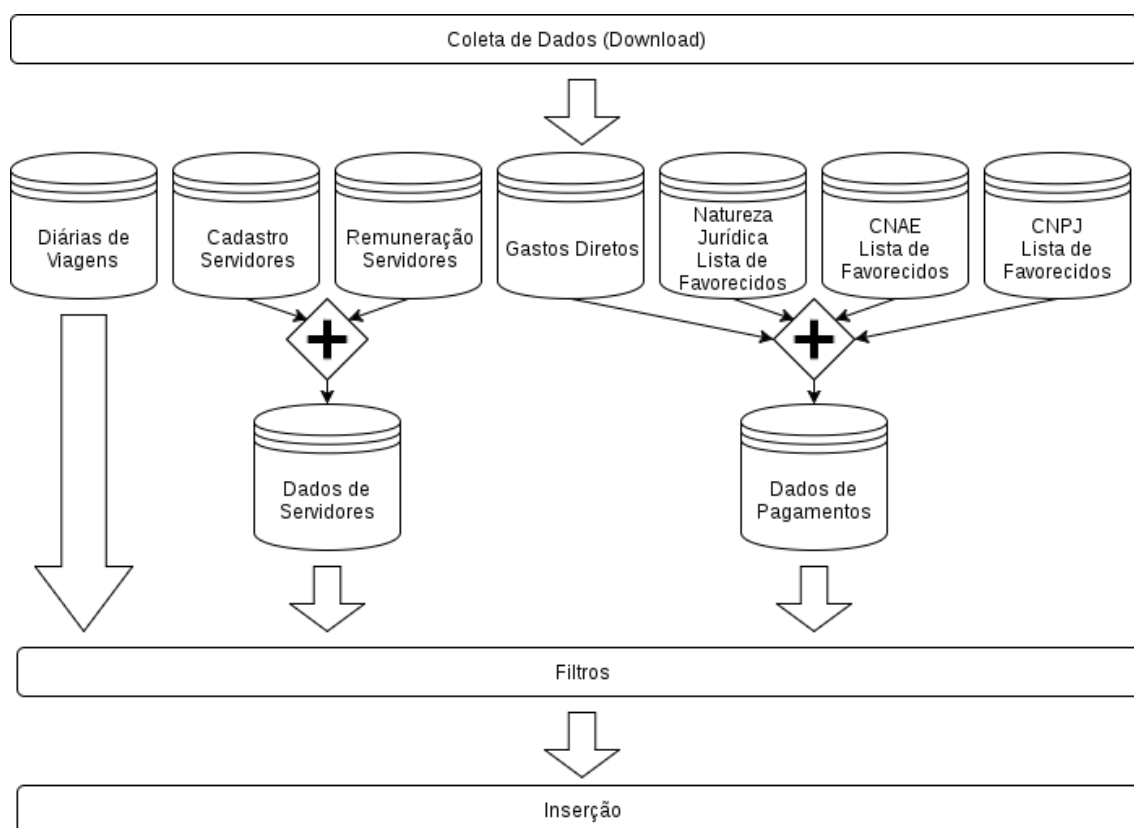


Figura 1. Visão geral do processo de coleta, tratamento e armazenamento dos dados do SIMTransparência.

quia, sociedade anônima, empresa individual, etc). **vi)** Lista de favorecidos por CNAE⁷, contendo nome de todas as atividades econômicas extraídas do cadastro da Classificação Nacional de Atividades Econômicas⁸. **vii)** Lista de favorecidos por CNPJ⁷, lista dos favorecidos - pessoa jurídica - dos pagamentos realizados pelos órgãos e entidades da Administração Pública Federal, que executam as despesas pelo Sistema Integrado de Administração Financeira do Governo Federal (Siafi).

Conforme exibido na Figura 1, as bases de “Natureza Jurídica”, “CNAE” e “CNPJ” são agrupadas em “Dados de Pagamentos”, com o objetivo de adicionar no conjunto de “Gastos Diretos” o setor de atuação e natureza jurídica das empresas favorecidas, informação que facilita a identificação das áreas de atuação (energia, administração, limpeza, etc) com gastos mais expressivos. A união entre “Cadastro Servidores” e “Remuneração Servidores”, por sua vez, permite visualizar dados pessoais dos servidores públicos (como nome, profissão e alocação) junto com sua respectiva remuneração. É importante salientar que a tabela de “Cadastro de Servidores” possui uma entrada para cada cargo ocupado pelo servidor (ex: professor e chefe de departamento) e, nesses casos, os dados de remuneração, que são disponibilizados como um único valor independente dos cargos ocupados pelo servidor, são preenchidos em apenas uma das entradas, com as demais recebendo valores nulos. Esse procedimento elimina problemas de dupla contagem em agregações que envolvam a remuneração de servidores.

⁸CNAE: <http://concla.ibge.gov.br/>.

A coleta é feita através de requisições *http* para o Portal Transparência do Governo Federal, que possui *urls* padronizadas de acordo com o conjunto de dados e data de disponibilização, com periodicidade mensal para a maior parte dos conjuntos. Como resultado, obtém-se arquivos *CSV*, que podem ser filtrados e armazenados de acordo com necessidade do usuário e capacidade computacional disponível.

3. Armazenamento

O armazenamento é feito no ElasticSearch, um banco de dados documental, de código aberto, distribuído e escalável, com capacidade de gerir grandes volumes de dados e fazer agregações de forma eficiente [Kononenko et al. 2014]. Ele tem como unidade básica um documento no formato JSON e documentos podem ser agrupados em conjuntos, formando índices.

O ElasticSearch permite a fragmentação dos dados dos índices em *shards*, que são distribuídos entre as máquinas do *cluster*, aumentando assim a eficiência das buscas e agregações. Por padrão, cada índice é separado em cinco *shards*, cada qual com uma réplica, totalizando dez fragmentos por índice. É importante considerar que a criação excessiva de índices implica em um alto custo computacional no gerenciamento dos fragmentos, ao passo que um número reduzido subutiliza o potencial de processamento paralelo e distribuído do ElasticSearch. No SIMTransparência, os índices são particionados pela fonte de dados (despesas, servidores e diárias), ministério (ex: MEC, MCTIC), além do mês e o ano em que os dados foram disponibilizados, gerando índices como *despesas-mec-2017-12* e *despesas-mctic-2014-03*. Acreditamos que esse modelo representa um crescimento controlado e com performance satisfatória.

4. Visualização

A visualização dos dados é feita com o Kibana, ferramenta desenvolvida para uso conjunto com o ElasticSearch. No Kibana, o administrador pode criar uma grande variedade de visualizações, desde gráficos de barra e pizza, até tabelas, mapas geográficos e de calor. Visualizações relacionadas compõem um *dashboard*, que pode ser compartilhado em páginas externas. O bom aproveitamento da ferramenta se dá através da criação de *dashboards* compreensivos formados por visualizações complementares, que permitam a exploração do mecanismo de filtros disponibilizado pela ferramenta.

Os filtros são aplicáveis em qualquer visualização através da seleção de um elemento de interesse, ação que propaga este mesmo filtro em todas as demais visualizações do *dashboard*. Tal mecanismo permite um número elevado de combinações que destaca, a cada interação, uma nova característica do conjunto de dados em análise. Por exemplo, no *dashboard* “MEC - Despesas” do SIMTransparência, podemos filtrar no gráfico “Descrição Seção” o elemento “Limpeza em Prédios e em Domicílios” e, em seguida, selecionar algum favorecido no gráfico ao lado, que já apresenta somente empresas relacionadas a limpeza. Mais acima, a tabela “Lista de Universidades” informa então quem possui o maior contrato com o favorecido selecionado. O usuário pode continuar a pesquisa habilitando novos filtros relacionados a data, elemento de despesa, ou outros, destacando situações de interesse em análise dinâmica. O repositório do SIMTransparência disponibiliza um arquivo com a definição dos elementos visuais e *dashboards* para serem carregados no Kibana (seção 5).

5. Instalação e Carregamento

Para replicar o SIMTransparência os seguintes passos são necessários: **i)** baixar o repositório do SIMTransparência, **ii)** opcionalmente instalar o ElasticStack (Logstash, Kibana e Elasticsearch), **iii)** executar programa que busca e carrega as bases de dados, **iv)** criar índices e **iv)** importar as visualizações e *dashboards* no Kibana.

O código fonte do SIMTransparência encontra-se no repositório GIT <https://gitlab.c3sl.ufpr.br/c3sl/transparencia.git>. Após baixar o repositório, siga as instruções presentes no arquivo README.md para criar exemplo contendo as bases de dados de despesas, diárias e servidores do Ministério da Educação. Em seguida, opcionalmente instale a suíte do ElasticStack da forma descrita na documentação e execute o comando abaixo⁹:

```
# ./simtransparencia --todas-as-bases --inicio 2017-01 --fim 2017-12
```

Após término da execução, crie os índices na interface gráfica do Kibana acessando-o pelo navegador (por padrão, em <http://localhost:5601>) e clicando no botão “Management” no menu esquerdo, em seguida em “Index Patterns” e “Create Index Patterns”. Insira então o valor `despesas-pagamentos-mec-*` e em “Time-field name” selecione Data Pagamento Timestamp. Repita o processo para criar os índices `despesas-diarias-mec-*` e `servidores-mec-*`. Por fim, para criar os gráficos e *dashboards* no Kibana, clique novamente em “Management”, “Saved Objects”, depois em “Import” no canto superior direito, selecionando por fim o arquivo `kibana_dashboards.json`, presente na raiz do repositório. Os *dashboards* podem agora ser acessados através do menu lateral esquerdo. Se preferível, é possível adicionar o parâmetro `--sem-elastic` e dispensar a ElasticStack. O programa irá gerar um arquivo CSV com os dados integrados que pode ser analisado com qualquer outra ferramenta.

6. Exemplos

Nesta seção será apresentado exemplo de visualização. Os dados são adquiridos sem modificação de valores a partir do Portal da Transparência do Governo Federal e a sua interpretação é de responsabilidade do leitor.

Um exemplo do gráfico de “Maiores favorecidos” pode ser visto na Figura 2. É notável que entre os dez maiores favorecidos, três são bancos. As instituições financeiras agem como intermediários em certas transações e, dessa forma, mascaram o destino final dos recursos. A mesma figura destaca também a participação da empresa “Sulclean” nas despesas da “Universidade Federal de Santa Maria” que, somadas as divisões de serviços gerais e segurança, corresponde a 25.61% do orçamento total da Universidade.

7. Desafios e limitações

Um dos desafios do SIMTransparência é disponibilizar os dados de toda a união, mas o volume de dados disponível requer uma grande quantidade de recursos computacionais. Essa limitação é atualmente mitigada pelo carregamento individual de apenas alguns ministérios, além da separação dos gráficos e tabelas em diferentes *dashboards*, o que reduz o número de consultas executadas no banco de dados a cada interação com o usuário. Nesse sentido, testes de performance e estabilidade são trabalhos futuros necessários para permitir o uso da ferramenta em larga escala através da interface publicamente acessível.

A interface do Kibana, apesar de proporcionar um grande avanço na visualização e navegabilidade se comparada ao atual Portal da Transparência, ainda possui diversas limitações.

⁹ # ./simtransparencia --help exibe ajuda sobre os argumentos de linha de comando.

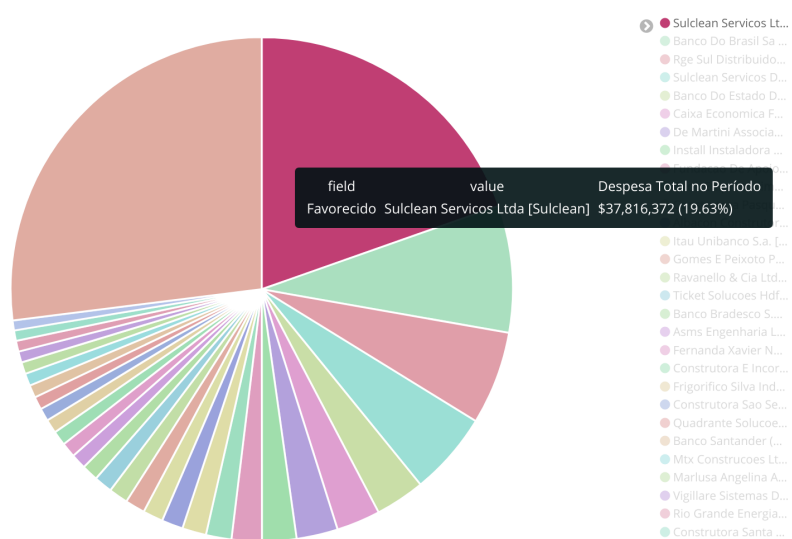


Figura 2. Gráfico exibindo maiores favorecidos com despesas na “Universidade Federal de Santa Maria”.

Ela não é internacionalizável, os filtros aplicados não estão sempre visíveis e, mais importante, operações de interjeição são limitadas a apenas alguns tipos de visualizações. O Elasticsearch, por sua vez, não permite junção entre diferentes índices e em determinadas situações¹⁰ calcula somente o valor aproximado das consultas, o que pode precisa ser tratado com cuidado em nosso caso de uso.

Por fim, os dados disponibilizados pelo Portal da Transparência possuem também limitações que, em parte, dificultam a análise. É possível observar, por exemplo, que bancos e fundações das Universidades são intermediários nas despesas da União, impossibilitando a identificação do destino final dos recursos aplicados. Ainda mais relevante, muitas entradas que são exibidas no Portal da Transparência não estão presentes nos arquivos CSV correspondentes disponibilizados para download, como o pagamento de mais de 458 bilhões em 2017 destinados a “Amortização e Juros da Dívida - Principal Corrigido da Dívida Mobiliária Refinanciada”¹¹.

8. Conclusão

Este artigo descreveu a combinação de uma série de bases de dados do Portal da Transparência do Governo Federal, com a geração de arquivos CSV e análise através da ferramenta gráfica Kibana. O SIMTransparência se coloca como uma visualização complementar aos dados disponibilizados no Portal da Transparência do Governo Federal, proporcionando algumas vantagens com relação a esse, como a liberdade para escolha do intervalo de tempo para análise e agrupamento de fontes diferentes em um único local. As diferentes visualizações disponibilizadas junto do repositório oficial permitem a gestores e usuários leigos a identificação rápida de despesas de maior relevância, proporcionando efetivamente a melhoria na transparência pública.

Referências

Kononenko, O., Baysal, O., Holmes, R., and Godfrey, M. W. (2014). Mining modern repositories with elasticsearch. In *Proceedings of the 11th Working Conference on Mining Software Repositories*, pages 328–331. ACM.

¹⁰Elasticsearch e as consultas de valor aproximado: <https://tinyurl.com/mh4yhb7>.

¹¹Link para o Portal da Transparência: <https://tinyurl.com/y7kcsavb>.

Uso de banco de dados orientado a grafos na detecção de fraudes nas cotas para exercício da atividade parlamentar

Gabriel M. Araujo¹, Maristela Holanda¹

¹Departamento de Ciência da Computação – Universidade de Brasília (UnB)
Brasília – DF – Brasil

gabrieljustware@gmail.com, mholanda@unb.br

Abstract. *This paper proposes the use of graph oriented databases to detect possible fraud in Quota for the Exercise of Parliamentary Activity, and the relationships between CEAP and open data from TSE regarding donations to brazilian elections in 2014. The use of these technologies, facilitates the manipulation of closely related data, both in terms of query complexity, and information visualization. The proposal in question was validated with a case study, using the open data of the Quota for the Exercise of the Parliamentary Activity of the Chamber of Deputies, and open data from TSE. An ETL was developed to extract the data and fill the database, then the queries were made to detect possible fraud and to obtain information about the data.*

Resumo. *Este artigo propõe o uso da tecnologia de banco de dados orientado a grafos para a detecção de possíveis fraudes na Cota para o Exercício da Atividade Parlamentar (CEAP), e o relacionamento da CEAP com dados abertos do TSE referentes as doações nas eleições de 2014. O uso dessas tecnologias facilita a manipulação de dados relacionados entre si, tanto em questão de complexidade na consulta, quanto em relação a visualização da informação. A proposta em questão foi validada com um estudo de caso, utilizando os dados abertos da Cota para o Exercício da Atividade Parlamentar da Câmara dos Deputados, e os dados abertos fornecidos pelo TSE. Foi desenvolvido um ETL para extrair os dados e popular o banco de dados, em seguida as consultas foram realizadas para detectar as possíveis fraudes e obter informações a respeito dos dados.*

1. Introdução

A política de transparência no Brasil surgiu há alguns anos em nossa sociedade, e tem como principal objetivo auxiliar na confiança da população sobre os serviços prestados pelo governo. O Brasil recentemente passou por um grande caso de corrupção, que foi a lava jato, e políticas de transparência também tem suas vertentes no combate a corrupção como é apresentado em [Diirr and Cappelli 2017]. O estudo feito por [Abramo 2000] compara as relações entre índices de percepção de corrupção e outros indicadores de alguns países latino americanos, e o Brasil se encontra na quadragésima nona posição em um ranking de corrupção dentre 90 países. Em [Filgueiras 2009] é feita uma pesquisa de opinião, na qual a Câmara dos Vereadores e a Câmara dos Deputados são as instituições com maior percepção de corrupção.

A análise feita por [Filgueiras 2009], direcionou a escolha do conjunto de dados utilizado neste artigo, que foi a Cota para o Exercício da Atividade Parlamentar (antiga

verba indenizatória). É uma cota única mensal destinada a custear os gastos dos deputados exclusivamente vinculados ao exercício da atividade parlamentar. O Ato da Mesa número 43 de 2009, detalha as regras para o uso da CEAP, entretanto um deputado pode realizar algumas transações que não são observadas facilmente pelos responsáveis em fiscalizar essas transações. Por exemplo, o artigo 4, parágrafo 13 do Ato da Mesa número 43 de 2009, diz: *”Não se admitirá a utilização da Cota para ressarcimento de despesas relativas a bens fornecidos ou serviços prestados por empresa ou entidade da qual o proprietário ou detentor de qualquer participação seja o Deputado ou parente seu até terceiro grau.”* Dessa forma, o Deputado pode realizar transações que violam essa regra, sendo inviável verificar as relações de parentesco de cada Deputado em cada transação, especialmente se utilizarem tecnologias inadequadas.

Em [Van Erven et al. 2017], é feito um estudo de caso utilizando um banco de dados orientado a grafos, que busca identificar fraudes em processos licitatórios no governo brasileiro. Nesse estudo de caso, é feita uma comparação entre a implementação utilizando um Sistema Gerenciador de Banco de dados (SGBD) orientado a grafos e um SGBD relacional. Essa comparação demonstra como as consultas em um banco de dados relacional ficam grandes e complexas, enquanto em um banco de dados orientado a grafos as consultas são menores, mais legíveis e fáceis de entender. Essa diferença está bastante relacionada com a estrutura que cada SGBD utiliza para representar os dados, de forma que o impacto vai além do tamanho e complexidade da consulta, mas abrange também a velocidade das consultas, bem como a facilidade de visualização das respostas.

Portanto, justifica-se o uso de um banco de dados orientado a grafo para identificar os relacionamentos envolvendo cada transação de um Deputado. Isso se deve porque os relacionamentos são evidenciados na estrutura de um grafo de forma muito mais natural e simples, onde cada entidade é representada como um nó do grafo e se relaciona com outras entidades por meio de arestas. Devido a essas particularidades, os bancos de dados em grafo vem ganhando bastante popularidade, tanto em pesquisas científicas, quanto em uso comercial, como mostra os trabalhos feitos por [Noble and Cook 2003] e [Chau and Faloutsos 2005].

O objetivo deste artigo é implementar um banco de dados baseado em grafo para evidenciar relacionamentos nas transações dos Deputados que violam o artigo 4, parágrafo 13 do Ato da Mesa número 43 de 2009, que regula a CEAP, e realizar o cruzamento com os dados abertos de doações das eleições de 2014, em busca de relacionamentos entre parlamentares e empresas. Para a implementação deste banco de dados, foi utilizado o SGBD NoSQL OrientDB. Para validar o modelo do banco de dados, algumas consultas de vínculos entre os deputados e as empresas são apresentadas.

Este artigo está dividido nas seguintes seções: a Seção 2 introduz o referencial teórico, abordando o tema de banco de dados orientado a grafos; a Seção 3 apresenta alguns trabalhos relacionados; na Seção 4 é apresentado em detalhes o desenvolvimento do projeto, e por fim, a Seção 5 apresenta as conclusões deste trabalho.

2. Banco de dados orientado a grafos

2.1 Definição de um grafo

O primeiro passo para entender um SGBD orientado a grafos é conhecer a estrutura de um grafo. A definição de um grafo pode ser feita da seguinte forma: um grafo G é uma tripla

ordenada $(V(G), E(G), \psi_g)$, que consiste de um conjunto não vazio $V(G)$ de vértices, um conjunto $E(G)$, disjunto do conjunto $V(G)$, de arestas, e uma função de incidência ψ_g que associa cada aresta de G um par não ordenado (não necessariamente distinto) de vértices de G . Dessa forma, se e é uma aresta e u e v são vértices, de tal modo que $\psi_g(e) = uv$, então, diz-se que e faz a união de u e v ; Os vértices u e v são chamados de extremidades de e [Bondy et al. 1976].

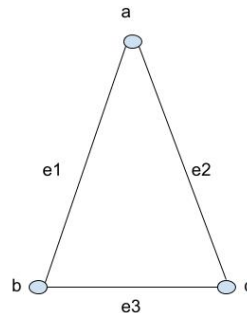


Figura 1. Exemplo de uma estrutura de grafo.

A Figura 1 apresenta um grafo, em que o conjunto $V(G)$ de vértices é não vazio e composto por três vértices, $V(G) = \{a, b, c\}$. Já o conjunto $E(G)$ de arestas é composto por três arestas, $E(G) = \{e1, e2, e3\}$. De forma que a função de incidência ψ_g é definida da seguinte maneira: $\psi_g(e1) = ab$, $\psi_g(e2) = ac$ e $\psi_g(e3) = bc$. Portanto, a definição formal do grafo acima é $(\{a, b, c\}, \{e1, e2, e3\}, \psi_g(e1) = ab, \psi_g(e2) = ac, \psi_g(e3) = bc)$.

2.2 SGBD NoSQL orientado a grafos

Os SGBD NoSQL orientado a grafos armazenam os dados em uma estrutura de grafo. Resumidamente, um grafo é um conjunto de nós e arestas, em que os nós são os objetos e as arestas representam o relacionamento entre dois objetos (nós). Esses SGBD usam a técnica conhecida como *index free adjacency*, em que cada nó possui um ponteiro diretamente para nós adjacentes. Essa técnica permite que uma grande quantidade de nós seja percorrida de forma eficiente [Nayak et al. 2013].

As consultas são feitas seguindo a ideia de percorrimento do grafo, o que torna esses SGBD mais eficientes que SGBD relacionais. O principal ponto de seu uso é em dados que são bastante relacionados entre si, pois a estrutura de um grafo expõe naturalmente os relacionamentos entre os objetos. O maior representante dessa categoria de SGBD é o Neo4j, que em sua versão livre não suporta a distribuição dos dados. Por esse motivo o OrientDB foi escolhido para a implementação do banco de dados deste artigo.

2.3 OrientDB

O OrientDB é um SGBD de código aberto sob a licença Apache. Ele é um SGBD NoSQL multi modelo, com suporte a uma arquitetura distribuída e orientado a grafos. Sendo assim

o OrientDB suporta operações com documentos, chave/valor e grafos. Essa característica garante flexibilidade para manipular os dados, sendo possível armazenar os dados tanto como grafos ou como documentos no mesmo banco de dados.

O OrientDB é implementado na linguagem java, tendo sua primeira versão disponível no ano de 2010. Ele possui alta flexibilidade para definir o esquema do banco de dados, podendo ser *Schema-free*, *Schema-hybrid* ou *Schema-full*. A sua linguagem de consulta é derivada do SQL e utiliza o modelo de transações ACID, isso demonstra que o OrientDB preserva a integridade dos dados ao mesmo tempo que também fornece um suporte a particionamento dos dados.

Todas essas características fazem com que o OrientDB seja um SGBD bastante flexível e confiável. As operações utilizando grafos em específico, vem ganhando bastante visibilidade pois funciona muito bem em certos domínios de aplicação. Como foi mencionado em [Gelbmann 2014] a popularidade dos SGBD orientados a grafos vem crescendo bastante nos últimos anos, e essas características ajudam a explicar o porque de banco de dados como o OrientDB e Neo4j estarem sendo utilizados em tantas aplicações.

3. Trabalhos relacionados

Na literatura, é possível encontrar alguns trabalhos acadêmicos onde banco de dados é utilizado para detecção de fraude. O trabalho desenvolvido por [Van Erven et al. 2017] propõe a modelagem para banco de dados em grafo utilizada neste artigo. Além disso, para validar a modelagem proposta é feito um estudo de caso para detecção de fraudes em processos licitatórios do governo brasileiro com o uso do Neo4J.

Outros trabalhos que utilizam estruturas de grafo como base para detecção de anomalias e fraudes incluem [Noble and Cook 2003] e [Chau and Faloutsos 2005]. O primeiro introduz duas técnicas para detecção de anomalias baseadas em grafos, enquanto o segundo busca detectar fraudes em leilões ao analisar o histórico de transações que existe no formato de um grafo.

Diferentemente dos trabalhos apresentados nesta seção, este artigo faz o uso do OrientDB como SGBD, e do conjuntos de dados da CEAP e do TSE, apontando o impacto do uso de estruturas de grafo na transparência.

4. Desenvolvimento

Esta seção descreve o processo de desenvolvimento desta pesquisa. Inicialmente, foram obtidos os dados abertos da CEAP no site da Câmara dos Deputados, e as receitas dos candidatos a deputado federal nas eleições de 2014 no site do TSE. Em seguida, foram feitas algumas transformações nos dados antes da fase de carregamento para o OrientDB. Foi desenvolvido um modelo de dados seguindo a modelagem proposta no trabalho de [Van Erven et al. 2018], que representa os dados persistidos no OrientDB. Em seguida, foram desenvolvidas as consultas para obter os relacionamentos entre os dados da CEAP e os dados das receitas nas eleições de 2014.

4.1 Dados Abertos

Foram utilizadas duas bases de dados abertos para o desenvolvimento deste trabalho. A primeira é referente aos dados da cota para exercício da atividade parlamentar, e pode ser

obtida no site da Câmara dos Deputados¹. A segunda base diz respeito as doações que cada deputado recebeu de empresas ou pessoas físicas, para a campanha eleitoral de 2014, e pode ser obtida no seguinte site do Tribunal Superior Eleitoral ².

A iniciativa de disponibilizar dados e informações por meio de portais tem como um dos objetivos melhorar a confiança da população nos serviços prestados pelo governo. A transparência governamental é, portanto, uma ótima iniciativa e extremamente benéfica para a população. Porém ainda existem diversos pontos a serem melhorados para que estudos sejam feitos de forma mais rápida e precisa.

Trabalhar com esses dados abertos se tornou um desafio, porque, cada instituição fornece os dados da própria maneira, essa falta de padronização dificultou bastante a fase de carregamento dos dados. Isso se deve ao fato de que a base de dados do TSE fornece o cpf de cada candidato, já a base da CEAP não fornece um identificador único para o deputado. Portanto, no momento de realizar o cruzamento não existia um identificador único em ambas as bases para fazer o relacionamento entre as bases de dados. Para resolver esse problema, se fez necessário realizar o cruzamento entre as bases por meio do nome de cada candidato, o que também foi um desafio, já que cada base utiliza um nome diferente para cada candidato. Os detalhes do processo de extração, transformação e carregamento serão fornecidos na Seção 4.3, o ponto principal a ser levantado é que trabalhar com dados abertos no Brasil, para realizar estudos e pesquisas, pode se tornar um desafio devido a falta de padronização entre as bases de dados.

4.2 Modelo de Dados

A modelagem dos dados seguiu o modelo GRAPHED [Van Erven et al. 2018]. A modelagem tem como objetivo dar uma visão geral de como os dados estão organizados no banco de dados, suas propriedades e relacionamentos. Dessa forma, o modelo de dados desenvolvido busca evidenciar as propriedades dos parlamentares, tais como: nome, partido e unidade federativa. Além disso apresenta características que identificam uma empresa como nome e CNPJ, e características atreladas a transação que o deputado faz com uma empresa como o valor e descrição da transação.

A Figura 2 apresenta a modelagem desenvolvida. como pode-se observar, foram propostas quatro classes que representam as instâncias dos vértices no banco de dados, que são: Parlamentar, Transação, Empresa fornecedora e Pessoa. As setas que saem de uma classe a outra, representa um relacionamento entre essas classes, que no grafo será representado por uma aresta. Dessa forma, uma instância de um parlamentar, realiza transações, que por sua vez está relacionada com uma empresa fornecedora. Esse caminho descrito, representa as transações da CEAP da Câmara dos Deputados. De forma análoga, uma empresa fornecedora realiza transações, que por sua vez favorece um certo parlamentar. Esse caminho representa os dados das doações das empresas para os deputados nas eleições de 2014, fornecido pelo TSE.

Os demais relacionamentos, como "socio-de" e "parente-de" tem por objetivo identificar possíveis fraudes na CEAP, uma vez que um parlamentar não pode utilizar a verba

¹<http://www2.camara.leg.br/transparencia/cota-para-exercicio-da-atividade-parlamentar/dados-abertos-cota-parlamentar>

²<http://www.tse.jus.br/eleitor-e-eleicoes/estatisticas/repositorio-de-dados-eleitorais-1/repositorio-de-dados-eleitorais>

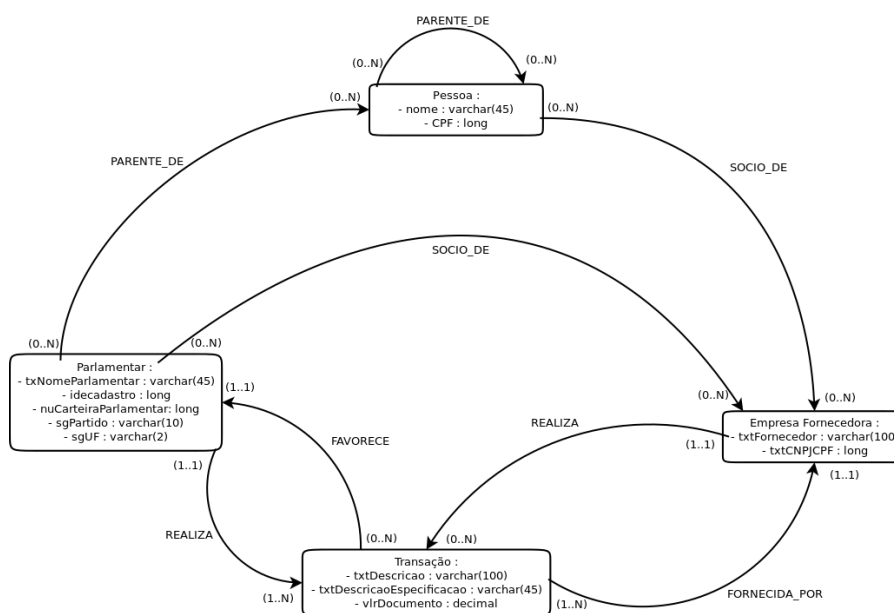


Figura 2. Modelo de dados seguindo o formato GRAPHED.

da CEAP com serviços de uma empresa que é sócio. Esses dois caminhos, se apresentaram como o maior desafio para o desenvolvimento do trabalho, uma vez que dados de parentesco dos parlamentares não são dados abertos.

4.3 ETL

Após a modelagem dos dados, foi desenvolvido um ETL (*Extract, Transform, Load*) utilizando a linguagem java, para extrair, transformar e carregar os dados para o OrientDB. O OrientDB foi desenvolvido na linguagem java, e executa portanto na JVM. Por esse motivo, o ETL foi feito em java, uma vez que, o OrientDB possui uma boa interface com essa linguagem. O ETL pode ser acessado por meio desse link no github³.

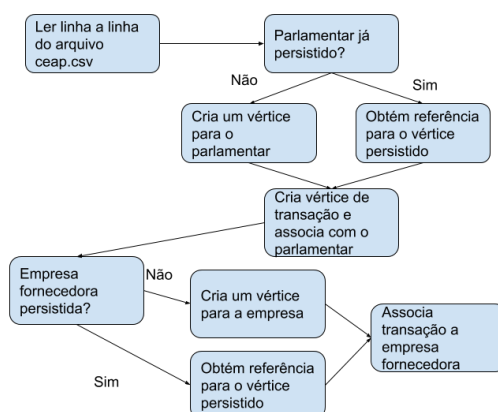


Figura 3. Fluxo de extração, transformação e carregamento para o OrientDB.

O fluxo do ETL é exemplificado na Figura 3. Dessa forma o fluxo se organiza em ler linha a linha do csv fornecido pela Câmara dos Deputados, e criar os vértices e arestas

³<https://github.com/gabrielmm1234/CEAP-ETL>

com base em certas condições. Por exemplo, não se pode ter mais de um vértice representando um parlamentar, mas no arquivo fornecido existem diversas transações para um mesmo deputado. O ponto chave desse processo são as associações entre os parlamentares com as transações e as empresas que forneceram o serviço. Criando esse vínculo, é possível por meio de consultas de casamento de padrões identificar possíveis fraudes ou padrões de transação e doação envolvendo um mesmo parlamentar e uma mesma empresa. A estrutura de um grafo fornece naturalmente os relacionamentos de forma simples e intuitiva, sendo uma ótima alternativa a estrutura relacional, que é muito comum atualmente. Além disso, a visualização da informação se torna mais clara e objetiva o que é uma ótima característica ao se tratar de transparência em dados governamentais.

O arquivo de dados da CEAP no ano de 2017 possui um total de 209496 transações, o carregamento dessas transações levou cerca de 16 horas em um notebook com Ubuntu 16-04 LTS, Intel Core i5-5200U CPU 2.20GHz * 4 e 6 Gb de RAM.

Já o arquivo das doações passou por um processo de filtragem, para obter somente os dados referentes aos deputados federais dos estados de Minas Gerais e Distrito Federal. Esse filtro para obter deputados federais dos dois estados foi feito para validar mais rapidamente a arquitetura e proposta de solução, no futuro demais estados serão adicionados no banco de dados. O arquivo já filtrado possui um total de 19302 doações de empresas a candidatura de diversos deputados. Claramente, somente alguns desses deputados foram de fato eleitos, e portanto deputados que não se encontravam na base da CEAP foram ignorados. O tempo total de processamento desse arquivo foi de 5 horas.

4.4 Resultados das consultas

Finalizado a obtenção e carregamento dos dados no OrientDB, foram feitas consultas em busca dos relacionamentos entre as empresas doadoras e os deputados de Minas Geras e Distrito Federal que utilizaram a CEAP com essas mesmas empresas. O OrientDB possui uma linguagem de consulta baseada na linguagem SQL.

A consulta é baseada no conceito de casamento de padrão, muito comum em linguagens funcionais. Dessa forma, para obter um certo padrão dentro do grafo o OrientDB fornece a função MATCH, como mostra a Consulta 1.

MATCH

```
{ class : Parlamentar , as : p } -RealizaTransacao ->
  { class : Transacao , as : t }
  -FornecidaPor -> { class : EmpresaFornecedora , as : e } ,
  { as : e } -RealizaTransacao -> { class : Transacao , as : t2 }
  -FornecidaPara -> { as : p }
```

RETURN \$elements

Listing 1. Consulta de relacionamento de doações entre deputados e empresas.

A Consulta 1 procura um padrão onde o mesmo é definido por parlamentares chamados de "p", que realizam transações da CEAP "t", fornecidas por uma certa empresa "e". Além disso essa mesma empresa "e", realiza uma transação de doação "t2", que é fornecida para o mesmo parlamentar "p" definido no início da sentença. O resultado dessa consulta nos deputados de Minas Gerais e Distrito federal é apresentado na Figura 4.

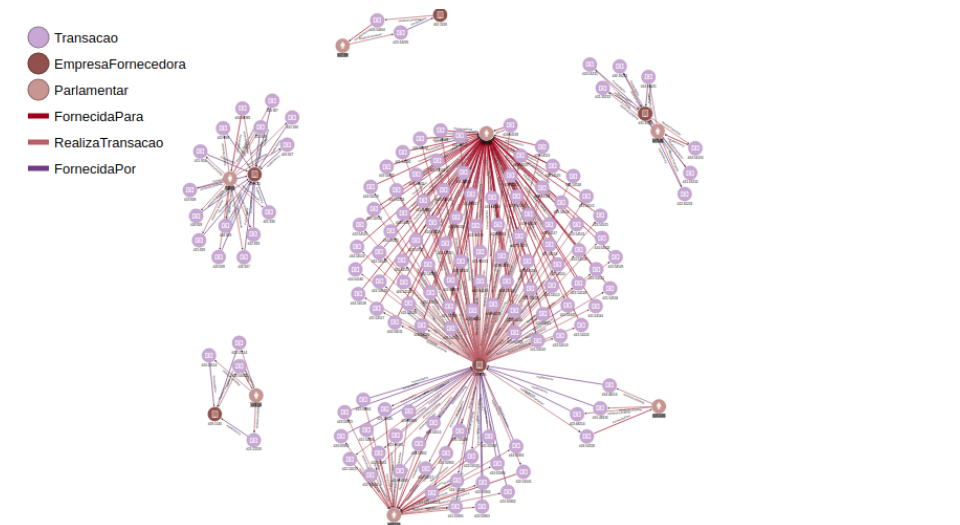


Figura 4. Resultado do cruzamento entre dados da CEAP e do TSE.

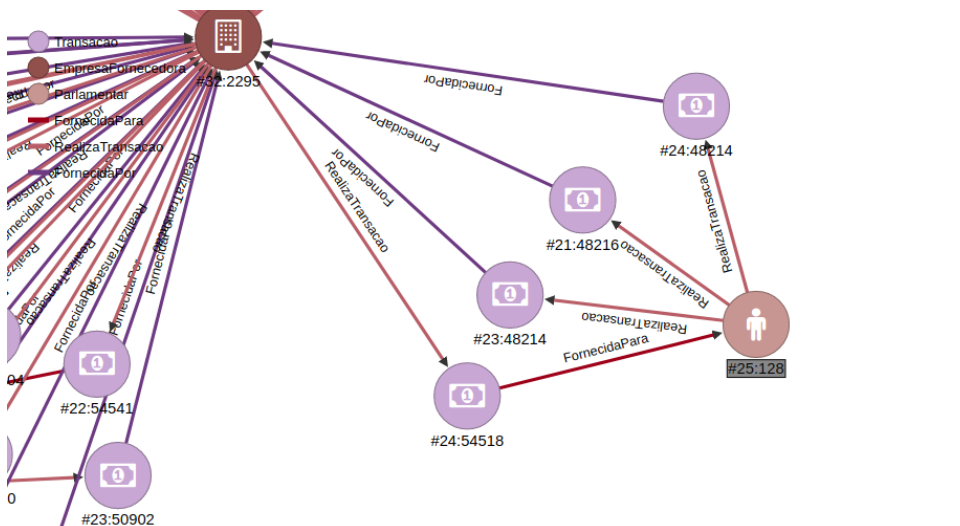


Figura 5. Padrão de relacionamento entre dados da CEAP e do TSE.

O padrão definido foi encontrado para um total de sete parlamentares todos em Minas Gerais. A Figura 5 mostra em mais detalhes como o padrão é definido. A partir da orientação das setas, é possível perceber que o parlamentar representado pelo vértice marrom claro e com o ícone de uma pessoa, realizou três transações, representadas pelo vértice roxo claro, com uma certa empresa representada pelo vértice marrom escuro. Sendo que essa empresa fez uma doação para esse parlamentar. No total, foram 44 transações registradas na CEAP, e 96 doações registradas pelo TSE que estão seguindo esse padrão.

Foi feita uma consulta para obter os padrões que podem ser considerados fraudes na CEAP, e testada com dados fictícios. No caso a consulta busca por deputados que usaram a CEAP com empresas das quais são sócios, como apresentada na consulta 2.

MATCH

```
{ class :Parlamentar , as :p } -RealizaTransacao ->
```

```

{ class : Transacao , as : t }
  -FornecidaPor -> { class : EmpresaFornecedora , as : e } ,
{ as : p } -Socio_De -> { as : e }
RETURN $elements

```

Listing 2. Consulta de relacionamento de uso da CEAP entre deputados e empresas nas quais o deputado é sócio.

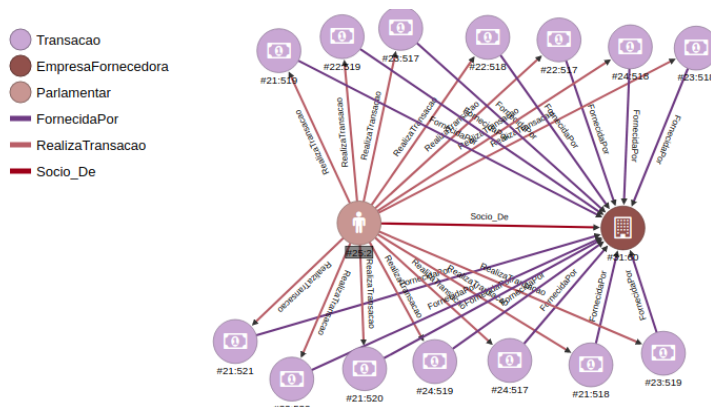


Figura 6. Padrão de uma transação fraudulenta com dados fictícios.

Como mostra a Figura 6, a consulta consegue localizar um padrão de transações efetuadas entre um deputado e uma empresa, na qual o deputado faz parte do quadro de sócios. A forma com que o grafo é apresentado visualmente, pode facilitar bastante o entendimento dos relacionamentos, de forma que fica claro que um deputado, representado pelo círculo com o ícone de uma pessoa, é sócio de uma empresa que prestou serviços com o dinheiro da CEAP. Dessa forma, a visualização gráfica, é uma das maiores contribuições de um SGBD orientado a grafos para a transparência, pois, como foi mencionado na Seção 1 a visualização dos vínculos em um grafo é mais simples e intuitiva do que uma estrutura de tabela utilizada nos SGBD relacionais.

5 Conclusões

Esse trabalho apresentou, portanto, o uso de banco de dados em grafo para detectar indícios de fraudes na cota para exercício parlamentar, bem como relacionamentos entre esses dados e os dados de doações para as campanhas eleitorais em 2014. Além disso, demonstra o desafio de se trabalhar com dados abertos no Brasil, uma vez que, os dados nem sempre são padronizados. Também mostra o uso dos bancos de dados NoSQL com dados abertos, e em aplicações que tem por objetivo apresentar informações relevantes a sociedade de forma objetiva e transparente.

Os resultados mostram como é intuitivo interpretar os dados olhando para a estrutura de um grafo, isso coloca os bancos de dados em grafos como ótima opção para aplicações que tem por objetivo a transparência e interpretação de dados complexos e bastante relacionados.

Para trabalhos futuros está incluso a implementação de um sistema colaborativo junto com a detecção de fraudes na CEAP. Este sistema irá utilizar o banco de dados desenvolvido neste trabalho, de forma a possibilitar que a população visualize diversas

informações a respeito da CEAP, e possa contribuir com informações importantes nas consultas que detectam as fraudes. Uma grande limitação deste artigo foi obter os dados de parentesco dos deputados, e o quadro societário das empresas, portanto, o principal objetivo do sistema é engajar a sociedade em cooperar com essas informações, e consequentemente superar essas limitações. Além disso, temos o uso de técnicas de aprendizagem de máquina nos dados da CEAP, que traz um problema interessante de executar essas técnicas em uma arquitetura baseada em um SGBD orientado a grafos, para encontrar padrões e criar modelos preditivos. Por fim, é importante realizar um estudo de como garantir a qualidade e veracidade dos dados, principalmente caso erros de software ou no processo de carga do banco de dados, cheguem a criar injustiças e acusar um deputado de fraude incorretamente.

Referências

- Abramo, C. W. (2000). Relações entre índices de percepção de corrupção e outros indicadores em onze países da América Latina. *SPECK, Bruno W. et al. Os custos da corrupção. Cadernos Adenauer*, (10):47–62.
- Bondy, J. A., Murty, U. S. R., et al. (1976). *Graph theory with applications*, volume 290. Citeseer.
- Chau, D. and Faloutsos, C. (2005). Fraud detection in electronic auction. In *European Web Mining Forum (EWMF 2005)*, page 87.
- Diirr, B. and Cappelli, C. (2017). Combate à corrupção através de relacionamentos interorganizacionais transparentes.
- Filgueiras, F. (2009). A tolerância à corrupção no Brasil: uma antinomia entre normas morais e prática social. *Opinião Pública*, 15(2):386–421.
- Gelbmann, M. (2014). Graph dbms are gaining in popularity faster than any other database category. https://db-engines.com/en/blog_post/26. Acessado em janeiro de 2018.
- Nayak, A., Poriya, A., and Poojary, D. (2013). Type of nosql databases and its comparison with relational databases. *International Journal of Applied Information Systems*, 5(4):16–19.
- Noble, C. C. and Cook, D. J. (2003). Graph-based anomaly detection. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 631–636. ACM.
- Van Erven, G., Silva, W., Carvalho, R., and Holanda, M. (2018). Graphed: A graph description diagram for graph databases. In Rocha, Á., Adeli, H., Reis, L. P., and Costanzo, S., editors, *Trends and Advances in Information Systems and Technologies*, pages 1141–1151, Cham. Springer International Publishing.
- Van Erven, G. C., Carvalho, R. N., de Holanda, M. T., and Ralha, C. (2017). Graph database: A case study for detecting fraud in acquisition of Brazilian government. In *Information Systems and Technologies (CISTI), 2017 12th Iberian Conference on*, pages 1–6. IEEE.

Transparência para Humanos e Máquinas: Um *framework* para Publicar Dados Abertos Interconectados Semanticamente Descritos

Julio Cesar dos Reis¹, Mario Ferreira de Brito¹

¹Instituto de Computação – Universidade Estadual de Campinas (UNICAMP)
Av. Albert Einstein, 1251, Cidade Universitária, CEP 13083-852, Campinas, SP – Brasil

jreis@ic.unicamp.br, profmariobrito@gmail.com

Abstract. *The Semantic Web allows data semantics to be explicitly described for humans and machines. Data transparency requires the structured publication of data for other systems. Open and interconnected data made available on the Web without copyright restrictions is key to achieving transparency mechanisms. However, often data to be published by organizations is found in many isolated systems that do not interact with each other. In this article, we propose a framework to enable the publication of open interconnected data, starting from several data sources for transparency. The study context is in a public university in Brazil where data of different nature need to become available to users of different profiles.*

Resumo. *A Web Semântica permite que a semântica dos dados seja descrita explicitamente para humanos e máquinas. Transparência de dados requer a publicação estruturada dos dados para outros sistemas. Dados abertos interconectados que são disponibilizados na Web sem restrições de “copyright” são a chave para atingirmos mecanismos de transparência. Mas muitas vezes os dados a serem publicados pela organização se encontram em diversos sistemas isolados. Neste artigo, propomos um framework para possibilitar a publicação de dados interconectados abertos, partindo de diversas fontes de dados visando a transparência. O contexto de estudo é em uma universidade pública onde dados de natureza distinta necessitam ser disponibilizados aos usuários de diferentes perfis.*

1. Introdução

A Web tem evoluído de documentos em hipertexto conectados para conceitos e dados interconectados que possibilitam integração de dados heterogêneos entre múltiplos sistemas de informação. Como evolução natural da Web, a Web Semântica investiga métodos visando a modelagem e descrição de dados com definições semânticas associadas [Berners-Lee 2001]. Nesse contexto, o conceito de *Linked Data* define uma série de princípios e técnicas para a publicação e o acesso de dados estruturados em repositórios Web [Wood 2014]. O principal objetivo é possibilitar a interligação de diversas fontes de dados e alavancar o uso e análise de dados.

Transparência refere-se a noção de tornar público informações de interesse da sociedade tornando visível informação a um ou mais interessados [Engiel 2015]. A Web se

tornou um ambiente propício para a publicação e consumo de dados abertos para promover transparência. Atualmente há diversos modelos de dados e princípios para a criação e disponibilização de dados abertos [Dadzie 2011].

Muitas iniciativas visando a transparência de dados públicos e governamentais têm adotado a Web como plataforma de compartilhamento. Essas iniciativas têm incentivado a disponibilização de dados abertos com o propósito de tornar o acesso aos dados sem restrições e transparentes [Janev 2012]. Esse movimento é principalmente observado com dados governamentais ou dados de pesquisa para permitir que a população ou outros usuários interessados tenham acesso transparente aos fatos das instituições e acompanhe a mesma conforme seus interesses.

No entanto, usualmente diferentes sistemas computacionais definem seus próprios modelo de dados heterogêneos em diversificados formatos, criando sérias dificuldades para a interoperabilidade e entendimento do significado dos dados por humanos e máquinas [Bizer 2009]. Esta pesquisa aborda a problemática de publicar dados abertos e semanticamente descritos com tecnologias da Web Semântica a partir de diversas fontes de dados. Os desafios envolvem o entendimento de modelo de dados de diferentes sistemas que codificam dados completamente distintos, de modo que seja possível entender o significado dos mesmos. O contexto prático de dados em universidade pública se caracteriza com sistemas computacionais dessa natureza.

Embora a literatura tenha abordado infraestruturas e aplicações para a publicação de dados interconectados abertos [Wood 2014], é necessário estudar técnicas que permitam estabelecer um processo desde a origem até a publicação, considerando a atualização dos dados periodicamente na Web. Isso envolve a criação de modelos para a descrição semântica e métodos para a geração e transformação automática de dados de um modelo para outro. Adicionalmente, os desafios envolvem a criação de endereços de identificação únicos e a criação automática de mapeamentos explícitos com conjunto de dados existentes. Em particular, mecanismos de atualização dos dados publicados precisam ser estudados para permitir consistência ao longo do tempo.

Este artigo investiga conceitos e técnicas para a publicação e atualização de dados interconectados abertos com semântica interpretável pela máquina. Propomos um *framework* computacional, nomeado de “*LODCamp*”, que permite gerar e atualizar dados interconectados para disponibilizar repositórios de dados estruturados e semanticamente descritos na Web visando a transparência de dados na universidade. O *framework* explora tecnologias da Web Semântica como RDF(s) e linguagens de descrição de ontologias OWL.

A metodologia de pesquisa envolve diversos passos iniciando-se pela revisão detalhada da literatura com o objetivo de entender os estudos e técnicas existentes. Isso também envolve uma análise minuciosa de infraestrutura e aplicações no contexto desta proposta. Em seguida, estudamos técnicas e modelos para a publicação e integração de dados abertos na Web Semântica.

Com base nesse contexto, e como nossa principal contribuição definimos um *framework* original para a publicação de dados abertos interconectados. Exploramos dados de uma universidade pública, a Universidade Estadual de Campinas (UNICAMP) em Campinas, São Paulo, para entender os desafios e prospectar soluções. Transparência é

relevante para a universidade e órgãos governamentais, e poucos trabalhos exploram tecnologias da Web Semântica para publicação de dados abertos nesse contexto, sendo esta a principal motivação para a condução dessa investigação.

O restante deste artigo está organizado da seguinte maneira: A seção 2 apresenta os conceitos fundamentais e trabalhos relacionados. A seção 3 descreve nosso *framework*, enquanto a seção 4 discute os resultados obtidos. A seção 5 finaliza o artigo com as considerações finais.

2. Trabalhos Relacionados

O formato heterogêneo de dados publicados na Web dificulta a interpretação desses dados e a ligação com outros conjuntos de dados disponíveis. Como meio de mitigar esse problema, propostas têm definido modelos de dados e técnicas que permitam a publicação e acesso a dados interconectados um com outros [Bizer 2009].

O modelo RDF baseia-se em identificar elemento de dados por meio de URI's, como identificadores únicos para os dados na Web, e descreve a relação entre dados via outras URI's. Essas são as chamadas triplas RDF (*Resource Description Framework*), em que se tem-se um sujeito, e um objeto que são ligados por um predicado. Com base no modelo de dados RDF, a proposta de “*Linked Data*” (dados ligados) [Bizer 2009] visa definir um conjunto de técnicas para a publicação de dados na Web, combinando e interligando diversos formatos e interfaces padrões com diversas fontes de diferentes localidades.

Berners-Lee *et al.* [Berners-Lee 2001] definiu um conjunto de princípios para a publicação de dados abertos na Web. Os princípios envolvem o uso de URIs para acesso global aos dados, uso de formatos não proprietários para a descrição dos dados, e sublinha a relevância da conexão entre elementos de dados distintos. De fato, o principal benefício é haver ligações explícitas e semanticamente definidas entre elementos de dados de maneira global, abertamente disponíveis na Web. Dadzie & Rowe [Dadzie 2011] discutiram os princípios de *Linked Data* na Web em grande escala e apontaram o grande interesse despertado pelas grandes empresas e entidades governamentais.

A semântica formalmente descrita e interpretável por máquina é obtida através do uso de artefatos como ontologias. Essas codificam o conhecimento interpretável por computador [Breitman 2007], tornando explícito para máquinas a semântica dos dados, que, por sua vez, facilita sua exploração nessa direção.

Janev *et al.* [Janev 2012] discutiram a evolução de dados abertos governamentais ao público em geral. Esse trabalho fez a extração de dados de mais de 100 servidores da Sérvia (antiga Iugoslávia) usando como ferramentas modelos RDFs e linguagem SPARQL¹, uma linguagem da Web Semântica que permite efetuar consultas em bases RDF. Nesse projeto, obteve-se a extração de preços, contas governamentais, usos de informações e tecnologias, assim também como atividades no geral realizadas pelo governo. Os autores concluíram que o modelo RDF trouxe benefícios para a publicações de dados nesse estudo de caso, permitindo uma grande quantidade de informação ser organizada e publicada de maneira interoperável na Web.

¹<https://www.w3.org/TR/rdf-sparql-query/>

Batista *et al.* [Batista 2013] argumentaram sobre a expansão da web semântica e usos das tecnologias RDF e OWL com *Linked Data* para divulgação dos dados. Apresentaram um estudo de caso sobre a publicação de dados abertos sobre os artigos do Simpósio Brasileiro de Banco de Dados. Definiram ontologias para a publicação do conjunto de dados permitiram a consulta aos mesmos via a linguagem SPARQL.

Bohm *et al.* [Böhm 2012] desenvolveram um sistema para coleta e disponibilização de dados abertos governamentais em larga escala. O sistema com interface Web permite usuários acessarem os dados e criar questões sobre políticos e fundos governamentais. As respostas são procuradas com base nos dados coletados e disponibilizados via formato RDF. Ainda nesse contexto, Moraes *et al.* apresentaram uma arquitetura para publicação e análise visual de dados interconectados usando padrão RDF visando aprimorar a gestão de dados na saúde pública.

Keßler & Kauppinen [Kessler 2015] indicaram que a publicação de dados abertos semanticamente descritos é essencial para a transparência na administração de Universidades. Eles definiram uma infraestrutura para a publicação de dados abertos da universidade de Münster, na Alemanha. Argumentaram que tal infraestrutura desenvolve um papel central para a visibilidade e acessibilidade dos dados produzidos e coletados pela universidade.

Fagundes & Sampaio [Fagundes 2017] discutiram a gestão pública administrativa e as dificuldades do cidadão no acesso as informações, tendo como base o portal da participação social “Participa.br”. Os autores indicaram que a sociedade brasileira está mudando de uma maneira a exigir dos órgãos governamentais, mais transparência em termos de gestão pública, possibilitando desse modo o acesso às informações pelo cidadão.

Silva *et al.* [Silva 2017] avaliaram o acesso às informações governamentais, em especial dos institutos federais da região centro oeste do Brasil. Os autores indicaram as dificuldades que brasileiros enfrentam no manuseio das informações disponibilizadas para visualização de dados governamentais. Concluem o trabalho mostrando através da pesquisa, que muitas unidades governamentais do centro oeste do Brasil não estão preparadas de acordo com a lei de transparência governamental reconhecida e aceita no Brasil.

Nossa análise da literatura indicou poucos trabalhos que visam explorar dados ligados descritos e publicados em RDF para integração de informação em universidades públicas. Nesta pesquisa, contribuímos com técnicas para a descrição, geração e atualização de dados heterogêneos descritos semanticamente via um *framework* de *software* para consumo de dados de diferentes fontes e geração automática de dados interconectados abertos. Acreditamos que nossa proposta inova no fato de permitir o usuário acessar informação organizada a partir de diferentes fontes. Exploramos dados da UNICAMP como estudo de caso e experimentação nesse trabalho.

3. Framework LODCamp

A figura 1 apresenta nossa proposta do *framework* onde os principais componentes envolvidos no sistema são destacados. Esse *framework* consiste de uma integração entre cinco módulos principais que são; A - as bases de dados da universidade (entrada principal de dados); B - módulo de acesso à vocabulários externos de domínios universitários; C - composição dos elementos de solução para a integração de dados (principais componen-

tes da solução); D - serviços e aplicações para consumo dos dados RDF; e E - base RDF para disponibilizar dados abertos (sistema Virtuoso para armazenamento de RDF).

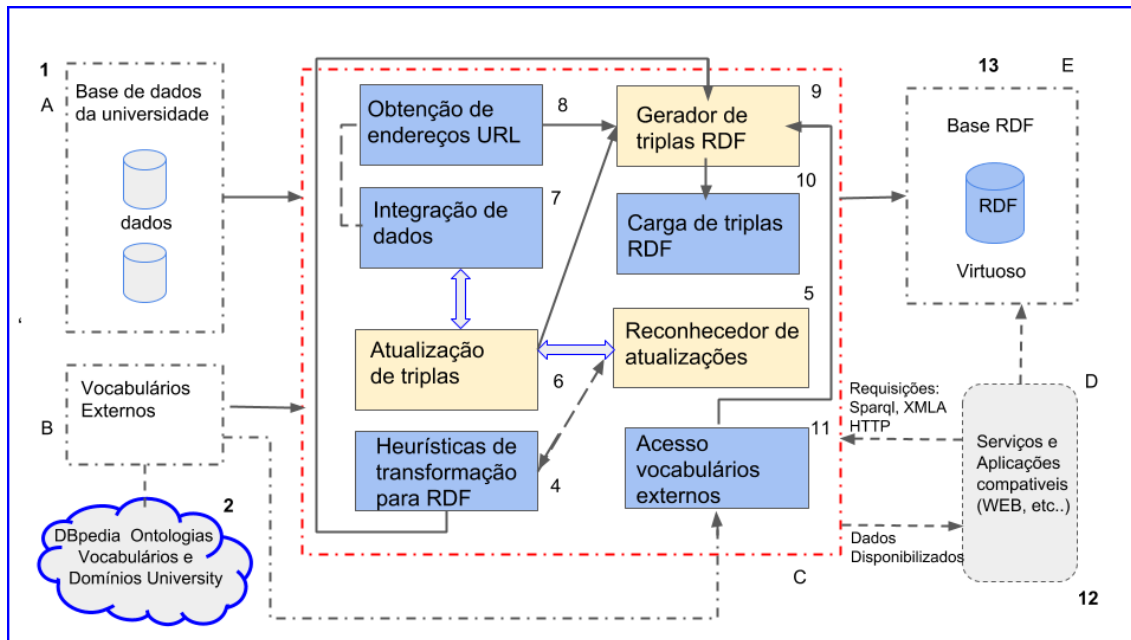


Figura 1. Componentes do *Framework LODCamp*

Inicialmente os dados da universidade se encontram disponíveis em diversas *Bases de dados da universidade* e outros tipos de arquivos que organizam dados. Esta é a principal fonte de dados para o *framework* e o início do processo de divulgação e transparência de dados. Consideramos como entrada arquivos que estão em diversos formatos tais como: .csc, .xml, .sql, .odf, etc., e devem ser tratados para ser disponibilizados aos usuários e sistemas finais.

Igualmente como entrada para o *Framework* estão os *Vocabulários externos* (B na Figura 1). O *LODCamp* deve acessar vocabulários estruturados que definem a semântica de dados do domínio. Para esse fim, consideramos ontologias sobre o domínio universitários e outras bases de conhecimento relevantes e ontologias que definem pessoas, e.g., *FOAF*². Por exemplo, a *DBpedia*³, como uma base de conhecimento que descreve artigos da Wikipédia, pode ser uma fonte para interligarmos dados da universidade com conjunto de dados externos, enriquecendo a divulgação. O *framework* define uma interface de software para permitir acesso à dados ligados disponíveis na Web de forma geral através de endereços URI e mecanismos de consulta de dados via SPARQL. A proposta considera o uso de vocabulários padrões já extensamente explorados e visa definir novas ontologias (usaremos a língua Portuguesa) para suprir as necessidades de publicação de dados na UNICAMP.

Os componentes do *LODCamp* (C na Figura 1) compõem os elementos para o tratamento dos dados. Esse módulo é responsável pela transformação dos dados de entrada em RDF. Inicialmente, O componente *reconhecedor de atualizações* (5) é responsável por

²<http://www.foaf-project.org/>

³<http://pt.dbpedia.org/>

monitorar a entrada de dados e reconhecer atualizações (novos dados ou dados modificados) advindos das fontes de dados da universidade. Uma vez reconhecida a atualização através do reconhecedor de atualizações, inicia-se o processo de atualização de triplas *Atualização de triplas* (6). Esse componente interage com o de Integração de Dados (7) e o Gerador de triplas RDF (9).

Uma grande problemática é efetuar e tratar a decisão entre triplas que representam elementos de dados existentes, e que devem ser modificadas, devido à alguma modificação no dado original, ou a criação de uma tripla totalmente nova, cujo dado não existia em formato RDF antes. A *Atualização de triplas* (6) é responsável por esta tarefa e a Integração de Dados (7) efetua a relação da tripla nova (atualizada) com outros elementos de dados. Por exemplo, um dado novo pode ser ligado a algum fato já existente na base RDF (interna ou externa).

O Gerador de triplas RDF (9) é responsável pela criação de fatos na base de conhecimento em RDF. Para esse fim, ele depende de três componentes chaves: *heurística de transformação para RDF* (4), Acesso à vocabulários externos (11) e obtenção de endereços URL (8). O componente de heurística é responsável por definir diretrizes de como as triplas serão definidas a partir de diversos formatos de dados de entrada. Nesse momento, é necessário o uso de vocabulários externos com as ontologias que definem as classes para representar a semântica dos dados.

A geração de uma nova tripla exige a definição de um endereço único para o item de dado. Para esse fim, o componente obtenção de endereços URI (8) é responsável por criar URI na geração de triplas. O sistema então verifica se houve acréscimo de arquivo ou se simplesmente foi substituído um arquivo existente. Caso seja um arquivo existente, o mesmo já possui um endereço URL e então este endereço é disponibilizado para o gerador de triplas. Caso contrário, caso seja um dado inteiramente novo, o sistema gera um novo endereço URI, disponibilizando-o e enviando uma informação ao gerador de triplas.

O componente de carga de triplas (10) efetua a transformação de fato dos dados em seu formato original para os fatos em RDF conforme a tripla gerada em (9) que são armazenados na base de triplas virtuoso⁴. As triplas conectam-se umas com as outras formando um grafo, e definindo os endereços URI de modo que possam ser acessados pelos sistemas finais de busca.

Os *serviços e aplicações* (D na Figura 1) são os sistemas de acesso as informações do módulo de integração de dados *LODCamp*. Nesse módulo, temos os navegadores Web e diversas requisições tais como HTTP e interfaces como o SPARQL para a requisição de dados no sistema.

Elaboramos um exemplo para ilustrar as etapas principais do *framework*. A Figura 2 apresenta um exemplo de dados de entrada na base inicial em uma planilha. Os dados brutos não indicam o significado dos itens e são necessários vocabulários externos para associar com os dados e possibilitar a descrição explícita da semântica.

O *framework* requer a criação de uma “URI”⁵ de base, nos quais os dados são

⁴<https://virtuoso.openlinksw.com/rdf/>

⁵Como exemplo ilustrativo, considere uma URI de base: `;;http://universidade/resources;ç;ç`. Usamos a sintaxe `@prefix uni: ;çhttp://universidade/resources;ç;ç` para se referir a esta URI.

	A	D	E	F	G	H	I	J	K
1	ALUNOS ATENDIDOS	2009	2010	2011	2012	2013	2014	2015	2016
2	Alunos atendidos com Bolsa Alimentação e Transporte	854	933	1.396	1.093	955	1.074	1.147	1.057
3	Alunos atendidos com Bolsa Alimentação e Transporte	30	30	30	30	38	32	30	30
4	Alunos atendidos com Bolsa Alimentação e Transporte	-	-	109	229	271	266	364	275
5	Alunos atendidos com Bolsa Auxílio Emergência	251	370	449	403	259	255	219	237
6	Alunos atendidos com Bolsa Auxílio Estudo Formação	-	-	-	-	-	30	78	69
7	Alunos atendidos com Bolsa Auxílio Instalação	-	-	-	-	111	193	112	200
8	Alunos atendidos com Bolsa Auxílio Intercâmbio	40	40	20	78	43	53	52	45
9	Alunos atendidos com Bolsa Auxílio Moradia	-	177	272	354	416	474	603	583
11	Alunos atendidos com Bolsa Auxílio Social	1.048	1.046	1.108	1.234	1.232	1.466	1.655	1.722

Figura 2. Dados iniciais de entrada (estruturados em uma planilha)

concatenados. Ela será usada como meio de identificar unicamente os elementos de dados que serão transformados em RDF.

O *framework* transforma a planilha de entrada em triplas RDF. Para isso exploramos a linguagem “Turtle”⁶ como meio de expressar as triplas RDF⁷. A Figura 3 apresenta duas triplas de exemplo que são geradas pelo *framework* com base na planilha de entrada. A Figura 3 (a) descreve que o sujeito “uni:bolsa_auxilio_moradia” é do tipo (uso da propriedade *rdfs:type*) “onto:bolsa”. Nesse caso, essa URI tem como prefixo “onto” que define uma ontologia do contexto universitário. Logo, “onto:bolsa” é uma URI da classe que representa o conceito de bolsa na universidade. Nesse contexto, a aplicação pode interpretar que “uni:bolsa_auxilio_moradia” é uma bolsa de estudo, e não uma bolsa, como utensílio de guardar objetos.

A Figura 3 (b) descreve o fato em uma tripla RDF em que “uni:bolsa_auxilio_moradia” está ligado via a propriedade “onto:ano_2010” ao valor 177. Nesse caso, estamos usando o @prefix *xsd* = “http://www.w3.org/2001/XMLSchema» para expressar que 177 é um valor decimal via o *schema xmls* já existente. A proposta permite reuso de vocabulários criados e disponíveis na Web, como esse exemplo de uso do vocabulário que expressa tipos de dados em XML.

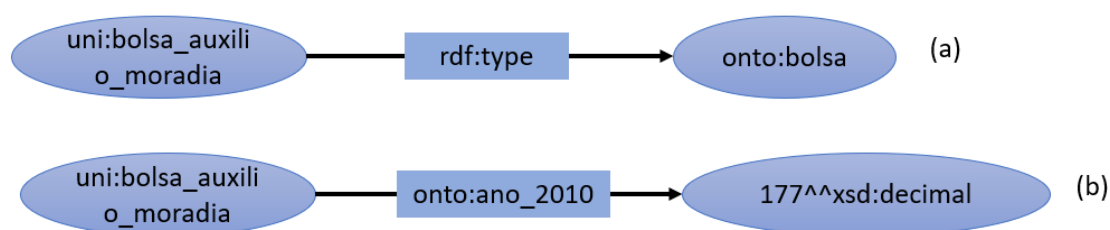


Figura 3. Exemplo de triplas RDF geradas com o uso da linguagem Turtle

Essas triplas são armazenadas na base de triplas RDF (virtuoso) para ser acessado por outras aplicações por meio de consultas e propostas de visualização de dados. Nosso exemplo objetivou apresentar uma visão geral das entradas e saídas do *framework* de modo a permitir acesso semântico aos dados da universidade. Elucidamos os passos de modo a comunicar o mecanismo de publicação de dados abertos através de princípios *Linked data* e uso de modelo de dados RDF.

⁶<https://www.w3.org/TR/turtle/>

⁷“Turtle” possui um formato simples em RDF para leitura final das pessoas

4. Discussão

Linked data é uma proposta que permite tornar explícito a semântica de dados publicados na Web e conseqüentemente alavanca a integração e recuperação de informação. Essa proposta desenvolve um papel chave em ser uma via para tornar mais acessível dados para pessoas e máquinas. É extremamente relevante tornar possível aos governos e administrações públicas a divulgação de dados públicos de uma maneira inovadora, de modo que usuários consigam acessá-los e usá-los. No entanto, argumentamos que simplesmente disponibilizar documentos na Web não significa que as instituições estão possibilitando transparência de dados. É preciso tratar os dados e disponibilizá-los de modo que usuários humanos e máquinas possam interpretá-los de uma maneira adequada.

Nesta pesquisa propomos uma infraestrutura como uma solução técnica para divulgar dados abertos semanticamente através de princípios de *Linked Data* e modelo de dados em RDF. Nossa inovação consistiu sobre a definição de um conjunto de componentes interligados no *framework LODCamp* que permitem a partir de dados de diferentes formatos e natureza serem publicados em uma base de dados RDF unificada. Em termos de transparência para humanos, entendemos que dados descritos em RDF não sejam apresentados diretamente para os usuários. Os dados brutos em RDF são adequados para a integração de dados entre sistemas, e o acesso dos usuários requer um mecanismo de visualização visando promover a transparência e entendimento facilitado dos dados pelo cidadão.

O *LODCamp* define componentes que atuam em todo o processo de publicação de dados, desde a chegada dos mesmos na base de dados da universidade, até sua saída para divulgação em formato RDF. O *LODCamp* atualiza os dados, efetuando a transformação e carregamento das triplas, entendendo as requisições, localizando os dados e disponibilizando-os aos sistemas finais para consultas. Trabalhos futuros envolvem investigar meios de permitir a visualização simplificada de triplas RDF.

Neste trabalho, analisamos o fluxo de divulgação de dados de uma maneira geral, desde a entrada de dados como fontes de dados da universidade (que é nosso caso) até a publicação final dos mesmos. Nosso *framework* pode ser base para construção de sistemas de *software* em outros tipos de órgãos governamentais para divulgar e atualizar ao longo do tempo a publicação de dados abertos na Web semanticamente descritos. Isso pode facilitar o acompanhamento pela população da administração pública governamental.

Como pontos que ainda necessitam ser aperfeiçoados nessa pesquisa são os mecanismos de monitoramento de atualização e as heurísticas de geração de RDF. Para esse fim, conduziremos observações em diversos dados na UNICAMP para entender os diferentes tipos e necessidades para atualização dos dados. Essa observação também beneficiará meios de formalizarmos diretrizes para as heurísticas.

Trabalhos futuros envolvem igualmente a definição de consultas e filtros adequados para acessos aos dados ligados integrados. Adicionalmente, prevemos desenvolver inteiramente o *framework* e conduzir estudos com dados reais da UNICAMP para avaliar a efetividade dos mecanismos de publicação e atualização de dados ao longo do tempo.

Acreditamos que explorar tecnologias já padronizadas da Web Semântica como RDF(S), SPARQL, Turtle, e plataformas existentes como o Virtuoso deva beneficiar o processo de publicação de dados abertos. Essas tecnologias desenvolvem um papel fun-

damental para a construção do *LODCamp*, fator esse relevante no acesso e divulgação dos dados abertos.

5. Conclusão

Transparência de dados em um contexto de diversidade em termos sociais e técnicos, como em uma universidade, demanda novas abordagens e suporte computacional para a publicação de dados que façam sentido para humanos e máquinas. Esta pesquisa contribuiu no desenvolvimento de técnicas para a geração de dados públicos e definidos semanticamente para pessoas e sistemas de *software* visando a transparência. Estudamos a construção de um *framework* que permita publicar dados interconectados abertos a partir de diversas fontes de dados heterogêneas advindos de diferentes setores da universidade. Descobrimos os elementos-chaves que devem compor uma solução para publicação de dados semânticos na universidade. Ilustramos o uso do *framework* abordando os desafios de entender e definir modelos comuns para a descrição semântica de dados de diferentes fontes de informação de uma universidade pública, a UNICAMP, através do *framework*. Trabalhos futuros envolvem o desenvolvimento de ferramentas de *software* que implementam os conceitos definidos nesta investigação.

Agradecimentos

Esta pesquisa tem apoio financeiro da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (Processo #2017/02325-5).

Referências

- Batista, Mateus Gondim Romão e Lóscio, B. F. (2013). Usando linked data para a publicação de dados abertos sobre o sbbd. In *XXVIII Simpósio Brasileiro de Banco de Dados, Recife, Pernambuco, Brasil*, pages 10:1–10–6.
- Berners-Lee, Tim e Hendler, J. e. L. O. (2001). The semantic web. *Scientific American*, 284(5):34–43.
- Bizer, C. e Heath, T. e. B.-L. T. (2009). Linked data - the story so far. *International Journal on Semantic Web and Information Systems*, 5(3):1–22.
- Breitman, Karin Koogan e Casanova, A. M. e. T. W. (2007). *Semantic Web: Concepts, Technologies and Applications*. Springer Verlag, London.
- Böhm, Christoph e Freitag, M. e. S. M. e. o. (2012). Govwild: Integrating open government data for transparency. Anais da Conferência WWW.
- Dadzie, Aba-Sah e Rowe, M. (2011). Approaches to visualising linked data: A survey. *Semantic Web*, 2(2):89–124.
- Engiel, Priscila, P. J. P. R. e. L. J. C. S. d. P. (2015). Representando o conceito de transparência através de um léxico ampliado da linguagem. WTrans - Workshop de Transparência em Sistemas.
- Fagundes, Vladimir e Sampaio, J. d. O. (2017). Análise da plataforma de participação social: Participação do cidadão e organizações na elaboração de políticas públicas. XXXVII Congresso da Sociedade Brasileira de Computação (5ºWTrans).

- Janev, Valentina e Milošević, U. e. o. (2012). Linked open data infrastructure for public sector information: Example from serbia. Anais da Conferência Internacional I-SEMANTICS.
- Kessler, Carsten e Kauppinen, T. (2015). Linked open data university of münster – infrastructure and applications. In *The Semantic Web: ESWC 2012 Satellite Events*, pages 447–451. Springer, Berlin, Heidelberg, Berlin, Heidelberg.
- Silva, Camila L. e Felberg, J. A. e. C. J. e. L. A. C. (2017). Um panorama sobre a acessibilidade para transparência em sites de institutos federais do centro-oeste do brasil. XXXVII Congresso da Sociedade Brasileira de Computação (5ºWTranS - Workshop de Transparência em Sistemas).
- Wood, David e Zaidman, M. e. R. L. (2014). *Linked Data: Structured Data on the Web*. Manning Publications CO.

A iniciativa da construção do Plano de Dados Abertos do Tribunal de Contas do Rio Grande do Norte através de parceria acadêmica com a Universidade Federal do Rio Grande do Norte

Frederico N. Pranto Filho¹, Renan Oliveira², Lindemberg Silva Pereira¹,
Joir Ramalho¹

¹Diretoria de Informática
Tribunal de Contas do Rio Grande do Norte (TCE-RN) – Natal – RN – Brasil

²Instituto Federal de Educação, Ciência e Tecnologia
do Rio Grande do Norte (IFRN) – João Câmara – RN – Brasil

fredericopranto@tce.rn.gov.br, renan.silva@ifrn.edu.br

{lindemberghsilva, joirramalho}@tce.rn.gov.br

Abstract. *Although there are several advantages and motivations for opening data by public government, there are also some difficulties and barriers to be overcome in order to achieve a Open Government Data (OGD) policy. Studies investigate the main motivations of developers and providers in working with OGD, as well as to identify the difficulties faced by them. The general objective of this work is to report the strategy adopted in the preliminary stages of the implementation of the Open Data Plan (ODP) of the Court of Auditors of Rio Grande do Norte (TCE/RN). The construction of the ODP tried to follow, as a working methodology, the directives defined in the ProceDA (Process for Open Data). In addition, we will discuss and share the gains obtained through a partnership between the TCE/RN and the Federal University of Rio Grande do Norte (UFRN).*

Resumo. *Embora haja várias vantagens e motivações para a abertura dos dados pelos órgãos da administração pública, há também algumas dificuldades e barreiras a serem superadas para a efetivação de uma política de Dados Abertos Governamentais (DAG). Estudos investigam as principais motivações dos desenvolvedores e fornecedores em trabalhar com DAG, bem como procuram identificar as dificuldades enfrentada por eles. O objetivo geral deste trabalho compreende relatar a estratégia adotada nos estágios preliminares da implantação do Plano de Dados Abertos (PDA) do Tribunal de Contas do Rio Grande do Norte (TCE/RN). A construção do PDA buscou seguir como metodologia de trabalho as diretivas definidas no ProceDA (Processo para Dados Abertos). Além disso, vamos discutir e compartilhar os ganhos obtidos a partir de uma parceria entre o TCE/RN e a Universidade Federal do Rio Grande do Norte (UFRN).*

1. Introdução

Uma publicação desenvolvida pelo Tribunal de Contas da União lista cinco motivos para a abertura de dados na Administração Pública [BRASIL 2015] e apresenta algumas razões

para que as organizações públicas invistam em iniciativas de abertura de dados governamentais. O TCE/RN, desde 2016, disponibilizou um Portal de Dados Abertos¹ em seu sítio contendo alguns conjuntos de dados no intuito de atender os cinco motivos citados anteriormente que são: (i) transparência na gestão pública; (ii) contribuição da sociedade com serviços inovadores ao cidadão; (iii) aprimoramento na qualidade dos dados governamentais; (iv) viabilização de novos negócios; (v) obrigatoriedade por lei. Dados abertos podem ser considerados como um dos importantes itens para o efetivo exercício do controle social. Controle social é a capacidade da sociedade de intervir nas políticas públicas com o objetivo de garantir direitos [Aguillar 1999].

A iniciativa de Dados Abertos possui diversos pontos positivos, como o potencial de gerar produtos e serviços úteis para a sociedade, fornecer mecanismos de transparência, fortalecer o controle social, a cidadania ativa, a cooperação e melhorias na administração pública além da criação de novas ferramentas, trazendo desenvolvimento e melhoria no bem-estar social [Kucera et al. 2016]. São apontadas ainda como áreas beneficiadas pela iniciativa o empoderamento do cidadão, o desenvolvimento de novos produtos e serviços, inovação, aumento na eficiência e efetividade de serviços governamentais, medição de impacto de políticas, e novos conhecimentos obtidos através da combinação de conjuntos de dados e de padrões em grandes volumes de dados [Open Knowledge International 2017]. Tudo isso é possível quando os dados disponibilizados possuem alto valor agregado [Janssen et al. 2012], possibilitando à própria sociedade explorá-los de maneira mais efetiva do que os próprios donos daqueles dados, que geralmente possuem apenas uma visão limitada sobre eles, utilizando-os para atividades básicas. Ao torná-los públicos, abre-se uma grande gama de possibilidades de utilização e visualização daquelas informações, trazendo benefícios em diversas áreas [Najdenov et al. 2015].

No período da publicação deste artigo, ainda tramita na Câmara dos Deputados o Projeto de Lei 7804/2014 [BRASIL 2014] que institui a Lei de Dados Abertos, estabelecendo o Comitê Gestor de Dados Público junto ao Ministério do Planejamento, responsável pela elaboração do Manual de Dados Abertos da Administração Pública e cria a obrigatoriedade para a disponibilização de dados abertos e de interfaces de aplicações web de forma organizada e estruturada para a União, Estados, o Distrito Federal e Municípios e dá outras providências. Atualmente, relacionado a Dados Abertos, existe o Decreto 8.777/2016 [BRASIL 2016] que institui a Política de Dados Abertos somente para o Poder Executivo Federal. Portanto, os órgãos das esferas estadual e municipal, assim como os Tribunais de Contas dos Estados e Municípios, ainda não são obrigados, por lei, a instituir o PDA ou o Portal de Dados Abertos. Entretanto, alguns portais já são uma realidade em alguns Tribunais de Contas^{2,3,4,5}, Estados⁶ ou Municípios⁷.

Os desafios que envolvem o ecossistema de dados abertos na administração pública já foi discutido em alguns trabalhos [Araújo 2017, Albano and Reinhard 2015, Fontoura 2015, Albano and da Silva Craveiro 2016,

¹Disponível em: <www.apidadosabertos.tce.rn.gov.br>

²Disponível em: <www.dados.tce.rs.gov.br>

³Disponível em: <www.tce.es.gov.br/portal-da-transparencia/dados-abertos>

⁴Disponível em: <www.portal.tce.pb.gov.br/dados-abertos-do-sagres-tcepb>

⁵Disponível em: <www.dadosabertos.tce.mg.gov.br>

⁶Disponível em: <www.dados.al.gov.br>

⁷Disponível em: <www.dados.natal.br>

Albano et al. 2017, Alcantara et al. 2015]. Embora haja várias vantagens e motivações para a abertura dos dados pelos órgãos da administração pública, há também algumas dificuldades e barreiras a serem superadas para a efetivação de uma política de Dados Abertos Governamentais (DAG). Segundo Araújo [2017], há vários fatores que motivam os desenvolvedores e fornecedores a trabalharem com DAG. O trabalho da autora procurou identificar também quais as principais dificuldades enfrentada e os pontos de melhorias no uso de DAG. Os fornecedores são os responsáveis por trabalharem na criação e fornecimento dos DAG. Já os desenvolvedores são usuários que possuem conhecimento técnico capaz de extrair, analisar e desenvolver aplicações que utilizam DAG.

O objetivo deste trabalho é compartilhar os ganhos obtidos a partir da parceria acadêmica entre o TCE/RN e a UFRN na iniciativa da construção do PDA do TCE/RN além de apresentar a estratégia e a metodologia utilizada para a publicação dos dados abertos.

2. A parceria entre o TCE/RN e o IMD

O Instituto Metrópole Digital (IMD) é uma Unidade Acadêmica Especializada da UFRN, cujo perfil alinha-se com a missão de fomentar a criação de um Polo Tecnológico em Tecnologia da Informação⁸ no estado do Rio Grande do Norte, abrangendo iniciativas dos setores público, privado e acadêmico.

O IMD atua na formação dos níveis técnico, superior e na pós-graduação. Dentre os cursos de pós-graduação há o Programa de Residência em Tecnologia da Informação aplicado à Área Jurídica e Órgãos de Controle, que inicialmente foi formalizado numa parceria entre o IMD e o Tribunal de Justiça do Rio Grande do Norte e na sequência em outros órgãos. A Residência promove o estudo e o aprofundamento de técnicas e metodologias de Tecnologia da Informação (TI) voltadas ao Poder Judiciário e Órgãos de Controle.

A parceria por meio da Residência entre a UFRN/IMD e o TCE/RN foi publicada no Extrato de Convênio N° 6/2017 [BRASIL 2017]. Os objetivos do convênio são, além da formação e aperfeiçoamento profissional, desenvolver sistemas de inovação tecnológica que irão contribuir na celeridade e qualificação do controle externo exercido pelo TCE/RN, promover avanços em áreas consideradas estratégicas, e estreitar os laços entre o desenvolvimento dos processos internos do órgão e o conhecimento acadêmico.

No total, dez servidores lotados na Diretoria de Informática (DIN) deste tribunal participam como discentes residentes no programa, sendo dois servidores vinculados na ênfase de Infraestrutura e o restante em Desenvolvimento e Business Intelligence (BI). Cada residente deve, como pré-requisito para a conclusão da residência, desenvolver um projeto de TI voltado a atender às necessidades do órgão. Dentre os projetos podemos citar soluções de BI, gestão de projetos, processo de desenvolvimento de software, melhorias em fluxo processuais e auditorias, política de segurança da informação, gerenciamento de acesso de redes sem fio e por fim, transparência. Este último é o tema norteador do nosso projeto.

Alguns projetos da Residência foram priorizados pelo Comitê de Tecnologia da

⁸Disponível em: <<https://portal.imd.ufrn.br/portal/parque-tecnologico>>

Informação do TCE/RN e outros puderam ser sugeridos pelos residentes. Este artigo apresenta o projeto de implantação do PDA no TCE/RN que tem como objetivos atualizar, reestruturar e expandir o Portal de Dados Aberto já existente no órgão e principalmente fomentar a cultura de dados abertos a fim de promover um dos pilares do Plano Estratégico dos Tribunais de Contas, o controle social.

Vale ressaltar que, durante o desenvolvimento deste trabalho, outras parcerias se concretizaram. Foram realizadas entrevistas técnicas com instituições que já haviam implantado seu PDA. Uma delas foi a UFRN, através da Superintendência de Informática, setor co-responsável pelo desenvolvimento do PDA na instituição de ensino, que sugeriu a utilização do ProceDA (Processo para Dados Abertos), processo este que lista grande parte das fases e tarefas realizadas por aquela instituição. Assim, após o alinhamento dos objetivos do nosso trabalho com o processo de publicação de dados abertos, decidiu-se pela utilização do ProceDA [Silva 2018] como norteador do nosso fluxo de trabalho. Além deste caso, por meio de mais um contato acadêmico na UFRN, fomos apresentados a uma solução de ETL (*Extract, Transform, Load*) integrada ao CKAN (*Comprehensive Knowledge Archive Network*) que está sendo desenvolvida por um grupo de pesquisa acadêmica [Macedo et al. 2017, Lopes et al. 2017].

Percebemos que, mesmo com o projeto ainda em desenvolvimento, há várias oportunidades para que outros órgãos possam consumir os dados custodiados no TCE/RN de forma estruturada. Até o presente momento realizamos contatos com secretarias dos governos do estado e setores da sociedade civil organizada. Acreditamos que a partir da promoção desta iniciativa, potencializamos o interesse de futuros consumidores de dados abertos após a futura publicação e divulgação do portal de dados abertos do TCE/RN. Algumas ações de promoção e divulgação do projeto foram realizadas por meio de *Workshops*, apresentação em eventos e discussões em redes sociais.

Conforme discutido em Araújo [2017], uma forma de garantir a motivação dos fornecedores de dados está ligada ao apoio que as universidades têm oferecido a esses atores. O caráter acadêmico da parceria e o interesse dos fornecedores na participação ativa de grupos de pesquisas envolvidas com a publicação ou o uso dos DAG se mostrou, conforme levantado pela autora, mais uma motivação para o desenvolvimento deste projeto.

3. O desenvolvimento do Plano de Dados Abertos no TCE/RN como metodologia de trabalho

Conforme discutido na seção 2, alguns projetos foram definidos e priorizados pelo órgão devido a sua importância estratégica. No entanto, além dos projetos pré-definidos, também foi possível a indicação de novos projetos para serem desenvolvidos na residência. Isto posto, foi criada e apresentada uma proposta para o desenvolvimento do PDA no TCE/RN e após o aceite do projeto pelo Comitê de Tecnologia de Informação do TCE/RN, foi iniciado o planejamento e o desenvolvimento do cronograma inicial de entregas do projeto. Na proposta, foi estabelecido que a metodologia de trabalho seria definida em um processo de publicação de dados, conforme será detalhado nesta seção.

A seguir detalharemos os seguintes pontos: (i) apresentação do processo de publicação de dados abertos; (ii) algumas adaptações do processo no nosso projeto; (iii) o PDA e a estratégia de seleção dos dados; (iv) e a apresentação da solução arquitetural

desenvolvida.

3.1. O processo de publicação de dados abertos

O Manual para Elaboração de Plano de Dados Abertos [BRASIL 2013] editado e publicado em 2013 pelo Ministério do Planejamento, Desenvolvimento e Gestão (MP) destaca as etapas que compõem o processo de construção do PDA. Para as demais tarefas relativas à publicação de dados, optamos por seguir as diretrizes apresentadas na proposta de Silva [2018], intitulada de ProceDA (Processo para Dados Abertos), por se tratar de um processo construído através da observação e compilação das melhores práticas presentes na literatura quanto ao processo de publicação de dados na administração pública. A Figura 1 ilustra suas fases, bem como sua sequência.

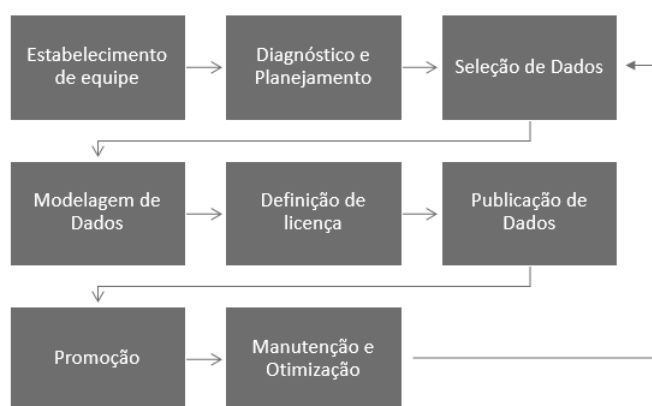


Figure 1. ProceDA e suas fases

O processo possui 8 fases, descritas a seguir:

- **Estabelecimento de Equipe:** definição da equipe que irá atuar no projeto de Dados Abertos da instituição, com o estabelecimento de papéis.
- **Diagnóstico e Planejamento:** consiste em avaliar o nível de preparo da instituição, realizar estudos de viabilidade do projeto, estabelecer um Plano de Dados Abertos (PDA), determinar a política de disponibilização de dados, estudo de legislação e outras questões gerais de projeto.
- **Seleção de Dados:** a instituição deverá selecionar os conjuntos de dados que serão publicados no ciclo atual, observando uma série de questões para o estabelecimento de prioridades, como a confidencialidade dos dados, o valor associado aos conjuntos de dados, benefícios e riscos da disponibilização de determinados dados, entre outros.
- **Modelagem de Dados:** deverá ser definido o formato de publicação dos dados, que pode ser: Dados Brutos, API ou Dados Conectados. A escolha deve ser baseada nas limitações de infraestrutura e o nível de valor agregado desejado para os dados.
- **Definição de Licença:** deve-se definir os termos de uso dos dados, que devem ser, resumidamente, de formato aberto, não proprietário, estável e de amplo uso. A publicação de um conjunto de dados sem estabelecer sua licença pode prejudicar sua utilização por parte de terceiros, indo de encontro aos ideais dos Dados Abertos.

- **Publicação de Dados:** são estabelecidas tarefas que devem ser realizadas ao se publicar os conjuntos de dados na Web.
- **Promoção:** são definidas estratégias de promoção dos dados que foram publicados, aumentando a chance de uso pela comunidade.
- **Manutenção e Otimização:** Deve-se avaliar a necessidade de atualizar conjuntos de dados previamente publicados, além de realizar monitoramento e ajuste das atividades realizadas durante todo o processo, buscando encontrar possíveis pontos de melhoria.

São previstos 5 papéis para a execução das diferentes tarefas associadas ao processo. Eles são descritos a seguir:

- **Órgão Central de Suporte e Controle das Ações (OCSCA):** órgão que ficará responsável por auxiliar e coordenar o processo de abertura de dados, removendo eventuais obstáculos e problemas que surgirem durante o caminho, bem como articulando todas as unidades que irão abrir seus dados.
- **Diretor:** Responsável pela iniciativa e principal interessado na abertura de dados. Geralmente é um Executivo da alta gestão, que pode ou não estar diretamente associado ao OCSCA.
- **Coordenador:** Responde pela execução do processo de abertura e acompanhará de perto as tarefas executadas.
- **Gestor Negocial:** Tem a capacidade de identificar dados importantes que podem ser abertos, os dados de risco e dados confidenciais.
- **Gestor Técnico:** Responde pelas ações técnicas do processo de abertura, como extração, elaboração de conjuntos de dados e sua apresentação.

Maiores informações sobre o ProceDA e um detalhamento da lista de tarefas de cada fase do processo, assim como os papéis responsáveis em executá-las, podem ser encontrados em Silva [2018].

3.2. Adaptações realizadas no processo

O processo definido no ProceDA sofreu algumas alterações no desenvolvimento das primeiras fases devido, principalmente, à forma com a qual este projeto foi idealizado. Como discutido no início deste artigo, o Decreto 8.777 institui a Política de Dados Abertos do Poder Executivo Federal. Desta forma, por força de decreto, os órgãos em questão tiveram de priorizar seus projetos internos a fim de atender o dispositivo legal. Acreditamos que a fase de Estabelecimento de Equipe, Diagnóstico e Planejamento e a primeira iteração da fase Seleção de Dados, provavelmente, puderam ser bem definidas nesses casos.

No nosso caso, a provocação inicial para o desenvolvimento do projeto não veio, necessariamente, por via legal ou por planejamento institucional. A iniciativa partiu da necessidade do desenvolvimento de um projeto acadêmico vinculado às necessidades do TCE/RN. Identificamos então uma oportunidade de fomentar a cultura dos dados abertos na instituição, além de contribuir para um dos pontos estratégicos definidos no Plano de Gestão (2016 - 2017) [ATRICON 2016] desenvolvido pela ATRICON (Associação dos Membros dos Tribunais de Contas do Brasil). Neste documento, um dos objetivos estratégicos apresentados é "Estimular a transparência das informações das decisões e da gestão dos Tribunais de Contas", que apresenta também duas iniciativas: (i) Apoiar

a instituição da Política Nacional de Fomento ao Controle Social e Dados Abertos pelos Tribunais de Contas; (ii) e apoiar a implantação ou o aprimoramento de ações destinadas a ampliar a transparência da gestão dos Tribunais de Contas. Assim, acreditamos que este projeto está alinhado com os objetivos estratégicos dos Tribunais de Contas.

Assim, algumas atividades das fases iniciais, como: a definição da equipe, estabelecimento de papéis e elaboração do PDA estão ainda sendo definidos de forma paralela às outras fases iterativas (ver Figura 1). Além disso nem todos os papéis foram completamente definidos até a publicação deste artigo.

3.3. O PDA e a estratégia de seleção dos dados

De acordo com o Manual para Elaboração de Plano de Dados Abertos [BRASIL 2013], o PDA é o documento que visa orientar as ações de implementação e promoção de abertura de dados, inclusive os geoespacializados, que deverão obedecer aos padrões mínimos de qualidade, de forma a facilitar o entendimento e a reutilização das informações. É ele quem organiza o planejamento referente à implantação e racionalização dos processos de publicação de dados abertos nas organizações públicas.

A implementação da Política de Dados Abertos, conforme estabelecido no Decreto 8.777, é realizada meio da elaboração e execução do PDA no âmbito de cada órgão ou entidade da administração pública federal, com isso, até a presente data, observamos a publicação de vários PDAs em diferentes órgãos. O processo de criação do PDA provoca a discussão do tema sobre transparência, dados abertos e controle social em cada órgão. Espera-se que a experiência do TCE/RN possa contribuir com outras iniciativas de organizações afins, conforme discutido em outros trabalhos [Bertin et al. 2017, Melo et al. 2017, Visoli et al. 2017].

O PDA está em desenvolvimento e será submetido à aprovação do Comitê de Tecnologia da Informação do TCE/RN para novas definições, prioridades e adequações.

A estrutura do PDA segue as recomendações do Manual para Elaboração de Plano de Dados Abertos e possui a seguinte estrutura:

- **Apresentação:** uma seção que explica e resume, de maneira clara e didática, o conteúdo e propósito do documento, com foco no cidadão.
- **Introdução:** informa que se trata da instituição do Plano de Dados Abertos do órgão, apresentando o normativo aplicável – incluindo padrões de organização e formatação dos dados a serem publicados - cenário institucional e, a seguir, os objetivos do documento.
- **Definição dos Dados a Serem Abertos:** nesta seção, enumera e apresenta os critérios para definição de que dados devem ser abertos e apresentar como foi feita a priorização.
- **Estratégias para Abertura:** detalha o passo-a-passo relacionado ao processo de publicação de dados, bem como premissas e requisitos de qualidade mínimos relacionados.
- **Catálogo no Portal Brasileiro de Dados Abertos:** aborda as responsabilidades relacionadas processo de publicação/catalogação de dados no Portal Brasileiro de Dados Abertos, incluindo o cadastro dos mantenedores dos metadados no portal.

- **Sustentação:** detalha as responsabilidades relacionadas à curadoria da publicação de dados e prever formas para garantir a disponibilidade dos dados abertos, bem como a incorporação do processo de abertura dentro da rotina do órgão.
- **Monitoramento e Controle:** apresenta o processo de monitoramento e controle das ações previstas no Plano, bem como relatório de prestação de contas.
- **Plano de Ação:** detalha as atividades e responsabilidades relacionadas ao processo de publicação de dados, bem como prazos e requisitos.

Com relação a definição dos dados a serem abertos, planejamos três principais entregas. A primeira entrega será disponibilizada na publicação do portal e tem como principais conjuntos de dados as informações institucionais divididos em cinco grupos: Contratos e Convênios; Despesa e Orçamento; Pessoas; Institucional; e Dados Auxiliares.

Na segunda e terceira entregas serão disponibilizados alguns dados enviados pelas Unidades Jurisdicionadas do TCE/RN, esses dados são recebidos e custodiados no tribunal. A segunda entrega tratará dos dados sobre a folha de pagamento e cadastro funcional dos servidores ativos, inativos e pensionistas dos jurisdicionados (SIAI DP) e a última entrega relacionados aos dados fiscais (SIAI Fiscal). Essas informações possuem um grande valor para a transparência, governança e controle social, visto que os todos estes dados estariam consolidados em um único local e em formato aberto.

3.4. Apresentação da solução arquitetural

A solução para a publicação dos dados abertos no Portal de Dados Abertos do TCE/RN que está sendo desenvolvido neste projeto está baseada no uso do CKAN⁹. O CKAN é uma ferramenta de catálogo de dados abertos, de código aberto, e desenvolvido pela *Open Knowledge Foundation*¹⁰, uma organização sem fins lucrativos que promove conhecimento livre. O CKAN é usado por governos nacionais e locais, instituições de pesquisa e outras organizações que coletam muitos dados. No Brasil, destaca-se o Portal Brasileiro de Dados Abertos¹¹, um portal disponibilizado pelo governo para que todos possam encontrar e utilizar os dados abertos e as informações públicas.

Para realizar a carga periódica dos dados no portal utilizamos a ferramenta ETL *Pentaho Data Integration*¹². A criação da infraestrutura do Pentaho foi realizada por outro projeto de BI da Residência de TI. Utilizamos a arquitetura já montada nos servidores do tribunal para criar e executar as transformações ETL dos conjuntos de dados. A carga é realizada diretamente para o CKAN através do *plug-in CKAN Resource Updater for Pentaho Data Integration (Kettle)*¹³.

4. Conclusão

O presente artigo apresentou a iniciativa, ainda em andamento, da construção do Plano de Dados Abertos do Tribunal de Contas do Rio Grande do Norte. Este projeto surgiu como pré-requisito de um curso de pós-graduação de nível lato sensu organizado a partir de um termo de cooperação entre o TCE/RN e a UFRN. O ProceDA (Processo para Dados

⁹Disponível em: <www.ckan.org >

¹⁰Disponível em: <www.okfn.org >

¹¹Disponível em: <www.dados.gov.br >

¹²Disponível em: <www.pentaho.com/product/data-integration >

¹³Disponível em: <github.com/localidata/CKAN-Resource-Updater-for-Pentaho-Data-Integration>

Abertos) está sendo utilizado como metodologia de construção do PDA. Além do PDA, o Portal de Dados Abertos está sendo reestruturado e atualizado por meio da plataforma de publicação de dados CKAN. Vale também ressaltar que as parcerias realizadas até o momento, seja com a academia ou com outros órgãos são fundamentais na aplicação dos conhecimentos acadêmicos nos processos da administração pública e mostram que esta relação é favorável para o desenvolvimento de soluções inovadoras.

Contudo, este artigo levanta somente alguns dados e processos iniciais da iniciativa da construção do PDA do TCE/RN, deixando em aberto os possíveis desafios na seleção e modelagem dos dados, as implicações sobre sigilo e privacidade de dados, validação da publicação dos dados no ambiente de produção, avanços e impacto na política de dados abertos governamentais do TCE/RN e estratégias de promoção dos dados abertos para a sociedade por meio de eventos e *hackathons*.

Como limitação do trabalho destacamos que embora outros tribunais já tenham publicado seus Portais de Dados Abertos não encontramos na literatura experiências sobre a publicação do PDA em outros Tribunais de Contas. Embora há estudos relacionado o nível de transparência dos Tribunais de Contas [da Silva et al. 2016, da Silva et al. 2014], ainda há a necessidade de uma investigação mais detalhada neste quesito.

Por fim, pretende-se com esta iniciativa, incentivar o desenvolvimento de um ecossistema de dados abertos no Estado do Rio Grande do Norte e nos Tribunais de Contas. Contribuindo assim para que essas ações fomentem a transparência e o controle social.

References

- Aguillar, F. H. F. (1999). *Controle social de serviços públicos*. Max Limonad.
- Albano, C. S., Araujo, M. H., and Reinhard, N. (2017). Fatores motivadores e facilitadores dos relacionamentos em redes: como os gestores públicos reconhecem esses fatores em dados governamentais abertos. *Navus-Revista de Gestão e Tecnologia*, 7(1):73–92.
- Albano, C. S. and da Silva Craveiro, G. (2016). Lições aprendidas com a utilização de dados orçamentários em formato aberto: Um estudo exploratório no ecossistema brasileiro. *Revista de Gestão e Projetos-GeP*, 6(3):17–27.
- Albano, C. S. and Reinhard, N. (2015). Desafios para governos e sociedade no ecossistema brasileiro de dados governamentais abertos (dga). *Cadernos Gestão Pública e Cidadania*, 20(67).
- Alcantara, W., Bandeira, J., Barbosa, A., Lima, A., Ávila, T., Bittencourt, I., and Isotani, S. (2015). Desafios no uso de dados abertos conectados na educação brasileira. In *Anais do Desafio Workshop de Desafios da Computação Aplicada à Educação*. CSBC.
- Araújo, N. M. (2017). Dados abertos do governo brasileiro: entendendo as perspectivas de fornecedores de dados e desenvolvedores de aplicações ao cidadão. Master's thesis, UFRN.
- ATRICON (2016). *Plano de Gestão (2016 - 2017)*. Brasília.
- Bertin, P. R. B., Machado, C. D., Visoli, M. C., Drucker, D. P., Pinto, D. M., et al. (2017). A construção do plano de dados abertos de uma organização pública de pesquisa e desenvolvimento e o desafio de uma ciência agropecuária aberta.

- BRASIL (2013). Manual para elaboração de plano de dados abertos.
- BRASIL (2014). *Projeto de Lei nº 7804/2014*. Brasília.
- BRASIL (2015). Cinco motivos para a abertura de dados na administração pública.
- BRASIL (2016). *Decreto nº 8777/2016*. Brasília.
- BRASIL (2017). *Extrato de Convênio nº 6/2017*. Brasília. Seção 3. Diário Oficial da União nº 204, 24 de out. 2017.
- da Silva, C. F., dos Santos, E. M. F., Chaves, M. C., Vaz, W., and Balaniuk, R. (2014). Dados abertos: uma estratégia para o aumento da transparência e modernização da gestão pública. *Revista do TCU*, (131):22–29.
- da Silva, W. G., Maciel, C., de Castilho, F. B. M., and Girata, N. N. H. (2016). Um método quantitativo para avaliar a adoção de dados abertos nos tribunais de contas do brasil. *iSys-Revista Brasileira de Sistemas de Informação*, 9(1):33–57.
- Fontoura, M. C. (2015). Hackeando dados abertos no brasil: motivações e práticas. *Cambiassu: Estudos em Comunicação*, 15(17).
- Janssen, M., Charalabidis, Y., and Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information Systems Management (ISM)*, 29(4):258–268.
- Kucera, J., Chlapek, D., Klímek, J., and Necasky, M. (2016). Methodologies and best practices for open data publication. In *Proceedings of the DATESO 215*, pages 52–64. CEUR-WS.
- Lopes, F., Barros, A., and Almeida, A. (2017). Processo automatizado de etl para dados abertos: Um estudo de caso do suap e ckan.
- Macedo, J., Cacho, N., and Lopes, F. (2017). A comparative study of tools for smart cities open data publication and management. In *Proceedings of the 2017 IEEE Summer School on Smart Cities (IEEE S3C 2017)*.
- Melo, J. d. S., Melis, M. F. M. d. S., et al. (2017). Análise da implementação da política de dados abertos no âmbito do poder executivo federal.
- Najdenov, B., Pejchinoski, H., Cieva, K., Jovanovik, M., and Trajanov, D. (2015). Open financial data from the macedonian stock exchange. In *Proceedings of the 6th Information and Communication Technologies Innovations 2014 conference*, pages 115–124. Springer Verlag.
- Open Knowledge International (2017). Open data handbook. Disponível em: <http://opendatahandbook.org/>. Acesso em 13 Out., 2017.
- Silva, R. O. (2018). Uma proposta de processo para implantação de dados abertos em instituições públicas brasileiras. Master's thesis, UFRN.
- Visoli, M. C., Bertin, P. R. B., Silva, A. R., Machado, C. R. D. L., Gonzales, L. E., Vacari, I., et al. (2017). Acesso aberto na embrapa: breve histórico, avanços recentes e desafios.

Desafios Sociotécnicos e Prospecções para Promover Transparência de Dados na Universidade

Breno Bernard Nicolau de França¹, Julio Cesar dos Reis¹, Rodolfo Jardim de Azevedo¹

¹Instituto de Computação – Universidade Estadual de Campinas (UNICAMP)
Av. Albert Einstein, 1251, Cidade Universitária, CEP 13083-852 – Campinas – SP – Brasil

{breno, jreis, rodolfo}@ic.unicamp.br

Abstract. *Data transparency is a key factor for the development of several sectors in society. Transparency not only boosts the understanding regarding what is being developed at public universities, but also shows how public resources have been invested. In this paper, we present results from a study on transparency, based on data collected from two initiatives for disseminating the transparency culture across the UNICAMP. We identified sociotechnical challenges, as well as described an architectural solution to improve transparency processes within the university to promote unrestricted access to public data.*

Resumo. *Transparência de dados é um aspecto chave para o desenvolvimento de diversos setores da sociedade. Em universidades públicas, a transparência potencializa o conhecimento sobre o que é desenvolvido permitindo o entendimento de onde os recursos são investidos. Neste artigo, apresentamos os resultados de um estudo sobre transparência, com base em dados coletados de duas iniciativas para difundir a cultura da transparência na UNICAMP. Identificamos desafios sociotécnicos, bem como apontamos uma solução arquitetural para facilitar os processos associados à transparência dentro da universidade e promover acesso irrestrito e facilitado aos dados públicos.*

1. Introdução

Acesso à informação para todos é essencial para o exercício da cidadania no século XXI. Pessoas necessitam de meios adequados para acessarem dados públicos visando obter informações claras sobre a eficiência das instituições. Isso se torna ainda mais relevante no contexto de instituições públicas, sobre as quais o cidadão demanda esclarecimentos sobre como os impostos recolhidos são aplicados. A transparência de dados desenvolve então um papel chave na garantia de governança pública adequada e combate à corrupção.

Dados governamentais abertos permitem melhor cooperação entre instituições públicas e o cidadão em geral. A abertura de dados não apenas beneficia a democracia por meio da transparência, como também permite melhor eficiência econômica com aprimoramento da gestão pública e disseminação de resultados [Mutuku and Colaco 2012]. Apesar dos benefícios teoricamente esperados, os desafios para a transparência demandam tanto investigações quanto mudanças organizacionais e tecnológicas para garantir acesso igualitário e público às informações governamentais [Jaeger and Bertot 2010].

A literatura tem apresentado diversas iniciativas para abertura de dados governamentais [Attard et al. 2015]. Mais especificamente, o uso de Tecnologias da Informação e Comunicação (TICs) tem indicado benefícios para promover transparência em dados e

umentar a clareza de ações governamentais [Bertot et al. 2010]. Particularmente, Bertot *et al.* [Bertot et al. 2010] mostraram que TICs desenvolvem um impacto relevante para criar mudanças sociais nas atitudes em relação à transparência. Nesse contexto, Böhm *et al.* [Böhm et al. 2012] apresentaram o *GovWILD*, um sistema web para integrar dados governamentais abertos, promovendo perguntas e respostas sobre políticos, campanhas e recursos governamentais.

Em particular, no contexto de universidades, Keßler & Kauppinen [Kessler and Kauppinen 2015] apresentaram uma infraestrutura para a publicação de dados abertos da universidade de Münster, na Alemanha. Eles indicaram que tal infraestrutura desenvolve um papel central para a visibilidade e acessibilidade dos dados produzidos e coletados pela universidade. Diversos projetos tem evidenciado a relevância da transparência e acesso a dados não apenas para gestão pública, mas igualmente para o desenvolvimento da ciência no geral, como a proposta de ciência aberta na medicina que visa a publicação e reuso de dados científicos [Krumholz et al. 2013].

Existem muitos dados de interesse público nos mais diferentes segmentos da sociedade, incluindo universidades públicas. No entanto, partes interessadas usualmente enfrentam dificuldade em encontrar os dados de interesse. Frequentemente, cidadãos buscam responder questões que demandam acesso e integração em fontes fragmentadas de informação. Esse problema é refletido pela utilização de diversos sistemas de informação isolados, que dificilmente interagem e permitem troca de dados de alguma natureza. O fato de não haver uma arquitetura para integração das informações, e que facilite o acesso ao usuário final com refinadas técnicas de consulta, cria barreiras na obtenção de respostas às necessidades de informação dos cidadãos.

Neste artigo, conduzimos um estudo sobre uma iniciativa de transparência de dados na Universidade Estadual de Campinas (UNICAMP). Analisamos os resultados obtidos em uma disciplina ofertada no contexto de graduação sobre transparência de dados e um evento de desenvolvimento de *software* promovido para obter ideias de soluções tecnológicas inovadoras visando transparência de dados na universidade.

Nossos resultados indicam desafios do ponto de vista organizacional e técnico. Dentre os desafios identificados, ressaltamos a falta de cultura sobre transparência tanto de usuários (no caso alunos) quanto de servidores públicos responsáveis por criar, organizar e disseminar os dados. Adicionalmente, o estudo realçou que transparência de dados requer inovações tecnológicas para disponibilizar efetivamente dados para um grande público-alvo. Assim, com base na análise, prospectamos uma arquitetura de *software* como uma infraestrutura para apoiar a disponibilização e acesso aos dados.

O restante deste artigo está organizado da seguinte maneira: a seção 2 apresenta a metodologia de pesquisa com os passos percorridos para obtenção e análise dos dados. Enquanto a seção 3 relata os desafios e a proposta de arquitetura, a seção 4 discute os resultados obtidos. A seção 5 apresenta as considerações finais.

2. Metodologia

2.1. Materiais e Participantes

Este estudo se fundamenta em duas iniciativas desenvolvidas na universidade para promover a cultura de transparência de dados.

Curso de transparência de dados. Uma disciplina motivada por alunos foi ofertada com o objetivo de discutir e entender aspectos, problemas e potenciais soluções sobre transparência na UNICAMP. Participaram, durante um semestre letivo, 51 estudantes de diversos cursos em nível da graduação. Os alunos foram então convidados a refletirem sobre questões de interesse em relação à universidade, dentro dos mais diversos aspectos.

Então, o curso foi organizado com base em três projetos desenvolvidos em grupos ao longo do semestre com os seguintes objetivos e atividades:

- **Projeto 1 (Aquecimento):** Definição de uma questão sobre transparência na universidade e tentativa inicial de respondê-la. O entregável desse projeto deveria ser no formato de um infográfico contendo a(s) pergunta(s) e resposta(s).
- **Projeto 2 (Investigar um tema):** Com base na sugestão de um conjunto de temas sobre diversos aspectos na universidade (*e.g.*, vivência no campus, orçamento, anuário estatístico, vestibular, comunidade, *etc.*), grupos definiram problemáticas que gostariam de investigar. Assim, os grupos deveriam relatar os passos seguidos para obtenção e entendimento dos dados, indicando os problemas encontrados.
- **Projeto 3 (Aprofundamento no tema):** Nesta fase, os grupos se aprofundaram em um aspecto particular do tema. Similar ao Projeto 2, os alunos relataram os meios e dificuldades em se obter e entender os dados sobre a universidade.

Todos os projetos seguiram uma lógica semelhante de entregas, sendo uma definição de problemas a serem investigados (por meio de questões ou temas pré-definidos), a coleta e disponibilização dos dados obtidos para responder às questões, e descrição do caminho (denominado em sala como “Odisseia”) utilizado para se obter às respostas, incluindo dificuldades/desafios. Particularmente, a “Odisseia” relatada pelos alunos foi utilizada para identificar os desafios descritos na Seção 3.1. Como última atividade, alunos foram convidados a criar um vídeo de aproximadamente 5 minutos apresentando os resultados descobertos durante o semestre.

Hackathon de transparência de dados. Neste evento, parcialmente organizado por alunos da UNICAMP, objetivou-se a proposição de soluções de *software* inovadoras com foco em transparência e dados abertos no contexto da universidade; no sentido de soluções para facilitar acesso, visualização e análise desses dados. O evento durou 36 horas e foi constituído de 13 times de alunos/desenvolvedores de *software* de diferentes instituições. Para contextualizar os alunos, diversas palestras com aspectos organizacionais e tecnológicos foram ministradas (*e.g.*, *Big Data* na Indústria, *Linked Data* na Prática, Transparência no Governo). A proposta dos grupos foram avaliadas por um comissão julgadora que deliberou 3 finalistas. Seção 3.2 apresenta os resultados e lições apreendidas.

2.2. Análise de dados

O seguinte procedimento foi conduzido para analisarmos qualitativamente os materiais coletados nos eventos descritos.

1. Exame detalhado do material obtido no curso (resultado do desenvolvimento de trabalhos dos alunos) para identificar problemas recorrentes de acesso aos dados da universidade e classificá-los;
2. Detecção e análise dos problemas endereçados e soluções propostas no *Hackathon* pelas equipes de trabalho;

O entendimento dos problemas em diversas perspectivas informou elementos relevantes para a concepção de componentes essenciais de uma arquitetura de software que permita melhor integração entre *software* existentes na universidade de onde os dados para transparência são providos.

3. Resultados

Apresentamos os desafios identificados no entendimento de transparência de dados na universidade dentro do contexto da disciplina ofertada (Seção 3.1), seguido das proposições obtidas e lições apreendidas no *Hackathon* (Seção 3.2). Como base nesses resultados, concebemos uma arquitetura de *software* para permitir a disponibilização integrada de dados oriundos de diferentes sistemas da universidade (Seção 3.3).

3.1. Desafios Identificados

O primeiro projeto da disciplina consistiu em levantar questões de interesse dos alunos a serem respondidas pelos mesmos através da disponibilização de dados públicos pela universidade. Exemplos de questões levantadas pelos alunos foram:

- *Como que chega o dinheiro na universidade?*
- *Quanto é o ganho e o gasto nos vestibulares?*
- *Quanto é o gasto mensal com bolsa para alunos?*
- *Qual a média de contratação de professores por ano na universidade?*
- *Quantos professores aposentados tem a universidade?*
- *Qual instituto recebe o maior repasse da universidade?*
- *Quantas pessoas almoçam nos restaurantes universitários diariamente?*
- *Qual o número médio de livros extraviados na universidade?*
- *Qual é o investimento mensal destinado para os ônibus internos?*
- *Onde são os pontos de Wi-Fi da universidade?*
- *Quantas vagas para deficientes físicos e idosos a universidade possui?*
- *Quanto é o gasto na segurança do Campus? Esse investimento quando aumentado, tem melhor resultado?*
- *Quantos alunos estrangeiros são matriculados em cursos da universidade?*
- *Qual o consumo médio de energia mensal da universidade?*
- *Quais são os patrimônios da universidade?*

Para os três projetos da disciplina, os empecilhos para chegar às informações foram capturados e classificados. Como eles foram descritos de maneira similar em todos os projetos e recaem sobre problemas comuns, sintetizamos os problemas nesta seção.

Considerando algumas questões e temas, sobretudo financeiros, as informações foram encontradas em menos de uma hora por meio de buscas *online* (via mecanismo de busca do *Google*), anuário estatístico da instituição, e sites institucionais. Nesses casos, mesmo que as informações pudessem ser completamente encontradas de maneira aberta, elas se encontravam em formatos distintos e de difícil manipulação (por exemplo, arquivos PDF e imagens), além de estarem frequentemente fragmentadas em diferentes fontes, necessitando de um esforço adicional para relacioná-las.

A grande maioria das questões e, portanto, os dados necessários para respondê-las precisaram de mais de três (chegando até seis) iterações, ou seja, idas e vindas em

diferentes setores da universidade, para se obter de fato os dados/informações adequados com a finalidade de alcançar as respostas corretas. Diante deste cenário, destacamos os seguintes desafios recorrentes:

Conteúdo de difícil entendimento sem conhecimento específico da área. Além do difícil acesso, os dados não são apresentados de forma adequada para o entendimento pela maior parte da população (leiga em conceitos específicos da universidade). Por exemplo, dados financeiros, questões orçamentárias, e políticas de distribuições de bolsas.

Temporalidade dos dados. Em qualquer caso dentre os observados, os dados refletiam apenas o passado e podiam demandar até 20 dias para obtê-los. Encontramos casos que as informações requisitadas em 2017, mas os alunos solicitantes receberam dados apenas referente ao ano de 2013.

Burocracia no acesso. A forma de acesso aos dados obtidos variou bastante dentro da mesma instituição. O acesso pode ser via informações estáticas abertamente disponíveis em portais da universidade, formulários de pedido de acesso a informação, *e-mails*, ofícios enviados por meio do Diretor de uma unidade, até presencialmente com reuniões agendadas. Informações sobre unidades, salários implementados na organização, puderam ser obtidos por meio de portais *Web*. Informações sobre questões acadêmicas como evasões, ingressantes em vestibular, informações socioeconômicas dos alunos, necessitaram de ofícios e foram recebidas em meios impressos.

Informações incompletas e/ou fragmentadas. As informações solicitadas por vezes estavam incompletas e, quando possível, o acesso à informação completa dependia do acesso à diferentes fontes, incluindo fontes externas à universidade, sendo umas *online* e outras obtidas por meio de ofício. Este foi o caso de informações solicitadas a respeito do patrimônio da universidade e investimento em projetos de pesquisa.

Indisponibilidade dos dados/informações. Algumas tentativas de se obter os dados solicitados foram respondidas, mas sem fornecer os dados/informações de fato. Nesses casos, foi alegado que o setor não possuía os dados ou que eles não estavam disponíveis e/ou permitidos para divulgação.

Conhecimento interno da organização para ter sucesso na solicitação. Existiram casos em que o solicitante precisou saber a qual setor interno da universidade ele deveria pedir os dados. Entretanto, o cidadão comum não tem esse conhecimento de como a universidade esta estruturada. É perfeitamente possível que uma solicitação realizada por um canal de propósito geral da universidade para acesso à informação tenha um resultado diferente da solicitação realizada por um canal específico de um setor interno. Isso evidencia limitações existentes na integração da universidade em relação aos dados e responsabilidades.

Contatos sem resposta. Algumas tentativas de fazer contato com os setores responsáveis simplesmente não foram respondidas. Esse foi um fator que dificultou o acesso à informação, mesmo que fosse necessário retirar pessoalmente algum documento.

Telefone de contato não alcançáveis. Em um caso específico, a única forma de contato encontrada com o setor (outra instituição pública detentora de dados da universidade) foi um número de telefone. Entretanto, o mesmo era bloqueado para ligações originadas de celulares.

3.2. Proposições e Lições Aprendidas no *Hackathon*

No geral, os grupos propuseram soluções diversificadas com objetivos específicos como detecção e visualização de anomalias na folha de pagamentos dos servidores, aplicação interativa com gráficos para apresentação de dados socioeconômicos e financeiros de cada instituto. Focamos na descrição dos objetivos e resultados dos três grupos finalistas avaliados pela comissão julgadora do evento. A motivação dos grupos foi, principalmente, o fato de haver dados de difícil acesso e análise, além de pouco intuitivos para usuários finais. De maneira geral, os grupos visaram disponibilizar os dados de maneira simples.

Visualização interativa de dados. Nesta proposta, criou-se uma solução de visualização interativa em um portal disponibilizando uma busca unificada para aprimorar a recuperação de documentos da universidade.

Chatbot. Esta proposta explorou o conceito de *chatbot*, como um assistente virtual de perguntas e respostas. O objetivo foi obter uma aplicação onde o usuário efetue perguntas sobre a universidade e o sistema de *chatbot* apresente resultados sobre a questão. Foi alcançado um protótipo de um *chatbot* implementado em uma rede social.

Centralização de dados. Esta proposta visou unificar os dados e disponibilizar seu acesso com técnicas de computação em nuvens através de uma única interface de programação. O objetivo foi permitir um aprimoramento na capacidade de cruzamento dos dados de naturezas diversas e facilitar a manipulação dos mesmos. No protótipo, um mapa interativo foi criado que disponibilizou relatórios de unidades da universidade com base nos dados centralizados.

O evento permitiu que os participantes refletissem sobre maneiras inovativas de disponibilizar os dados para um grande público-alvo. Tecnicamente, o evento permitiu aos alunos descobrirem o uso de novas tecnologias e linguagens de programação explorando o tema transparência de dados.

3.3. Proposta de Arquitetura de Software

Os resultados obtidos indicaram que a transparência requer o acesso e análise de dados armazenados em sistemas computacionais distintos. Esses dados estão dispersos em diversas fontes, são representados em modelos de dados distintos, com formatos heterogêneos, e em sistemas implementados com linguagens e por equipes distintas. Adicionalmente, os dados descrevem informação de diversos domínios (vida acadêmica de alunos, vestibular, orçamentário/financeiro, funcionários e docentes, entre outras), sendo por vezes desconexos e não pré-processados para permitir análise e entendimento dos mesmos. Assim, propomos uma arquitetura computacional para permitir o desenvolvimento organizado de uma infraestrutura para disponibilização e acesso aberto aos dados (*cf.* Figura 1).

O componente “*Backoffice*” é acessado pelas unidades da universidade, as quais disponibilizam os dados para acesso, por meio das interfaces desse componente (“*CarregarDocumento*, *CarregarImagem*, e *CarregarDados*”). Ele permite que ambos dados estruturados (*e.g.*, CSV, XML, JSON, SQL) e artefatos como documentos PDF e imagens sejam inseridos na plataforma, por vias distintas. É importante ressaltar que essas interfaces devem estar disponíveis de forma a serem acessadas por usuários e integradas com outros sistemas da universidade. O gerenciador de transformação de dados (“*GerenciadorTransformacoes*”) e o gerenciador de artefatos (“*GerenciadorArtefatos*”) são

responsáveis por essas tarefas. Ao carregar quaisquer informações ao “*Backoffice*” por meio de suas interfaces, metadados devem descrever a unidade da organização detentora dos dados (origem do dado), bem como os domínios nos quais os dados se aplicam (planejamento, financeiro, pesquisa, ensino, *etc.*).

A arquitetura prevê dois tipos de repositórios. Um para armazenar artefatos como documentos e imagens, por exemplo editais, regulamentações, fotografias e plantas de obras, e outro repositório focado na centralização e relação entre dados. O repositório de dados estruturados necessita considerar modelos compartilhados de dados e permitir uma descrição flexível dos dados para possibilitar análises e descoberta de relações entre os dados de diversos domínios. Base de dados orientada a grafos, ou ainda bancos de dados não relacionais de alto desempenho em consultas, podem acomodar esses requisitos e permitir consultas com alto desempenho e disponibilidade.

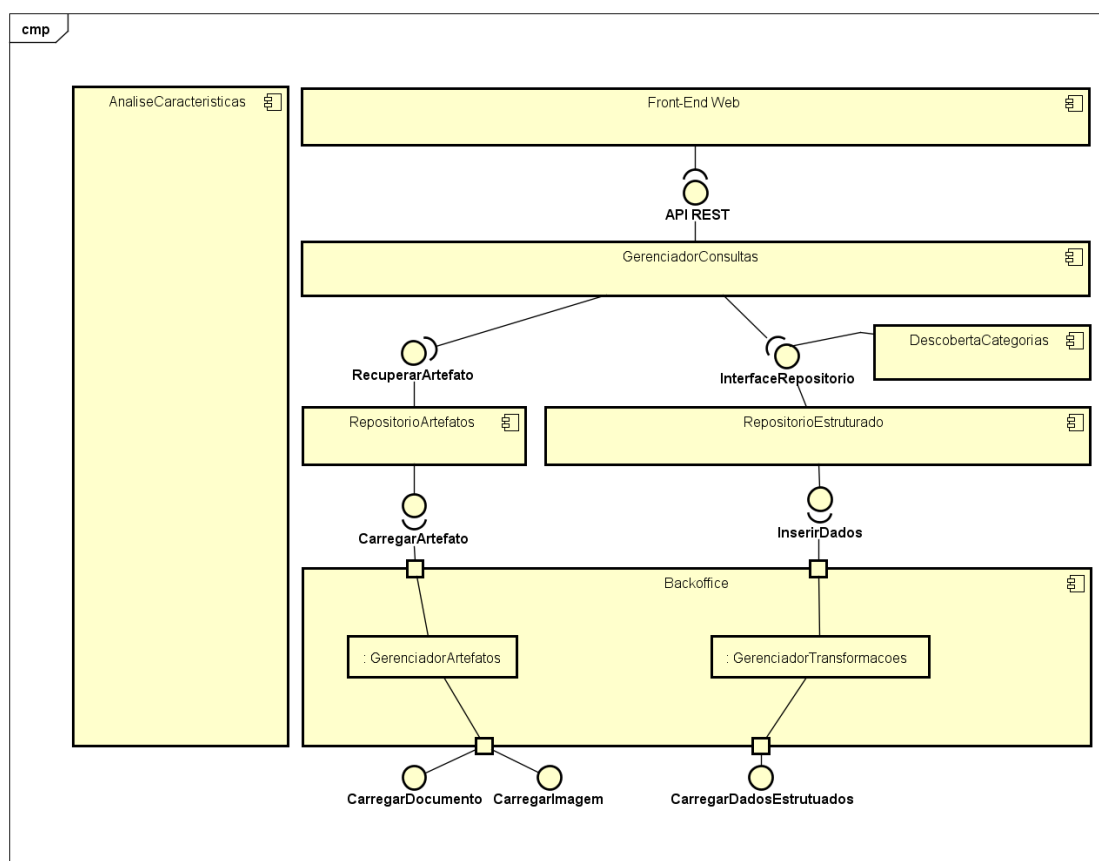
O componente “*GerenciadorConsultas*” é responsável por tratar consultas em ambos repositórios. Consultas podem ser expressas por meio de diferentes facetas e filtros, além de palavras-chaves e categorias de análise pré-determinadas. O componente de descoberta de categorias (“*DescobertaCategorias*”) visa extrair possíveis classes de consulta de interesse do usuário através de técnicas de classificação e agrupamento aplicados sobre os dados do repositório e de consultas submetidas pelos usuários.

O componente de “*Front-End Web*” visa implementar interfaces Web com usuários finais para expressar consultas a dados e múltiplas perspectivas de visualização desses dados (gráficos, tabulações, entre outras). Ainda, é relevante considerar o perfil de acesso e utilização dos usuários de forma a entender como eles utilizam e que tipos de informações são mais acessadas por determinados perfis. Para esse fim, o componente “*AnaliseCaracteristicas*” prevê capturar os dados mais acessados, as visualizações mais exploradas, os domínios de maior interesse dos usuários, de forma a direcionar a atenção para determinados conteúdos ou funcionalidades que necessitem de melhorias.

4. Discussão

O interesse em conhecer melhor a universidade, entendendo como os investimentos são realizados e qual retorno se obtém desses investimentos, perpassa as diferentes áreas da universidade, desde situação orçamentária e infraestrutura até elementos do dia-a-dia no campus. Esse interesse se dá em função do desconhecimento sobre o funcionamento interno da universidade e sua importância para a sociedade, até certo ponto justificáveis pelo tamanho e complexidade da instituição.

O problema caracterizado nos desafios para o acesso aos dados da universidade possui questões técnicas e sociais, do ponto de vista político-organizacional e da cultura da mesma. Do ponto de vista social, por exemplo, é interessante notar que certos dados são negados ou dados como indisponíveis por um setor em seu canal imediato com o cidadão, mas são disponibilizados se contatar as pessoas “certas” dentro desse setor. Isso evidencia ou a falta de conhecimento entre os funcionários de determinados setores sobre quais informações eles detêm e quais deveriam estar disponíveis ao público, ou um receio sobre os riscos associados à disponibilização de determinadas informações ao público. Na prática, nota-se haver uma desconfiança, quando se solicita informações, de que alguém fará mal uso das informações ou que poderá prejudicar o indivíduo ou o setor responsável pela informação. Isso justifica, em parte, uma quantidade razoável de negativas ao tentar



powered by Astah

Figura 1. Proposta de Arquitetura

solicitar informações de caráter público, mas que o livre acesso é desconhecido pelas pessoas internas àquelas organizações.

Ainda explorando os aspectos não-técnicos, percebe-se que não há uma cultura de transparência, no sentido de estabelecer quais, por quem e como os dados devem ser disponibilizados à sociedade. Internamente nos órgãos da universidade, parece haver uma alta proteção sobre esses dados, muitas vezes sem fundamento, onde qualquer informação apenas deve ser disponibilizada mediante autorização direta de cargos da alta gestão da universidade como coordenadores, diretores, pró-reitores e reitores.

Percebemos, também, que a utilização de determinados jargões ou termos no ato da solicitação pode influenciar no mal entendimento sobre quais dados se deseja obter, ou ainda recair na situação do setor entender que aquele dado não está sob sua responsabilidade. Um exemplo é quando se pede informações de cunho financeiro a setores/unidades que não são responsáveis pela questão orçamentária da universidade. Nesses casos, as repostas tendem a redirecionar o solicitante a outros setores ou simplesmente responder negativamente à solicitação, mesmo sendo informações específicas daquele setor.

Não é razoável assumir que o cidadão necessite entender a estrutura interna da instituição pública para requerer dados. A transparência deve seguir no sentido inverso, fornecendo um canal de comunicação único com o cidadão. Esse tipo de serviço já é disponibilizado pela UNICAMP. Entretanto, não se pode afirmar que tal serviço funciona de

forma completamente transparente ou facilita a obtenção dos dados, uma vez que grande parte das informações necessitam ser pedidas via formulários.

Entendemos que as iniciativas de transparência ou dados abertos de instituições públicas deveriam, idealmente, ser pró-ativas no sentido de entender as demandas da sociedade e disponibilizar as informações sem a necessidade de solicitações via formulários, ofícios, e outros meios mais burocráticos. Entretanto, a universidade passa por um momento de entendimento desses aspectos e a transição no processo de transparência necessita, também, do apoio tecnológico.

Diferentes órgãos públicos fazem uso de portais de dados abertos, os quais funcionam como uma ferramenta de busca, e repositório de dados sobre diferentes áreas da esfera pública. Entretanto, a proposta desses portais assume que o cidadão conheça a estrutura interna do governo, entendendo por exemplo, quais dados estão associados a que setores. Este estudo detectou algo similar na universidade, onde é necessário conhecer posição geográfica dos institutos ou órgãos formais da instituição para se obter informação, por vezes de forma presencial.

Os aspectos técnicos endereçados na proposta da arquitetura supõem um esforço reduzido das unidades detentoras dos dados para disponibilizá-los, evitando um custo considerável em uma eventual integração sistêmica entre os diversos sistemas utilizados na universidade. Assim, a arquitetura prevê não somente meios para disponibilização dos dados, mas componentes que permitam o melhor entendimento das demandas do cidadão em relação aos dados públicos da universidade, por meio de componentes de análise de dados de uso e disponibilizados nos repositórios. O grande desafio da arquitetura é lidar e definir um modelo compartilhado de dados para permitir consultas mais complexas.

Identificar os problemas existentes e encaminhamentos de soluções de transparência para a universidade tem sido uma preocupação recorrente. Esforços como este estudo podem ajudar no diagnóstico de problemas, ainda que em situações particulares e com a limitação em termos de abrangência das soluções propostas. Alguns dos desafios identificados podem ser entendidos como de percepção geral, mas que nessa oportunidade ficam evidenciados, seja por meio dos relatos das “Odisséias” dos alunos ou pelas inúmeras solicitações formais (ofícios, formulários e *e-mails*) de acesso aos dados. Ainda, entendemos que iniciativas desse tipo disseminam a importância do tema, ajudam a criar uma cultura de transparência na universidade, e desenvolve um papel educacional para com os alunos no entendimento do tema e de aspectos relacionados à universidade.

Do ponto de vista dos alunos, conseguimos notar um maior engajamento e curiosidade à medida em que mais dados eram coletados pelos grupos e apresentados em sala de aula. Enquanto, no início, a maior parte dos alunos estava preocupada com questões orçamentárias, o foco foi sendo transferido para outras questões como resultados acadêmicos (evasão, alunos por curso, *etc.*), uso de recursos pelos hospitais, patrimônios da universidade, organização de setores, *etc.* O *Hackathon* trouxe outra dinâmica ao trabalho, agregando código ao processo e fomentando ligações entre os dados já coletados. Ao final do semestre, tanto a visão sobre a universidade quanto a capacidade de analisar os dados encontrados foram aprimoradas.

Este estudo evidenciou características chaves do problema, mas entendemos que não é possível utilizá-las como um diagnóstico completo do estado da universidade em

relação ao tema de transparência. Os casos enfrentados representam situações particulares, mas que se repetem em diferentes níveis e setores da universidade. Ainda, não temos como afirmar que outras universidades, ou ainda outros órgãos públicos do país enfrentam necessariamente os mesmos problemas em função da observação e coleta de dados localizada. Finalmente, a solução arquitetural encaminhada deve ser amplamente discutida com os setores de TICs da universidade para evoluí-la e entender desafios específicos.

5. Considerações Finais

Transparência de dados na universidade envolve tanto desafios organizacionais quanto tecnológicos com inovações que permitam desenvolver soluções que façam sentido aos usuários finais. Neste artigo, conduzimos um estudo para entender os desafios sociotécnicos para a criação de soluções visando transparência de dados no contexto universitário. Nosso estudo detectou diversas dificuldades para encontrar informações sobre a universidade. Propomos uma arquitetura como meio de organizar o desenvolvimento de *software* de uma plataforma de transparência na universidade. A arquitetura permitirá que dados de diferentes naturezas sejam integrados para a resposta de consultas e filtro de dados relevantes. Trabalhos futuros envolvem investigar meios de implementar e implantar a arquitetura, utilizando-a como peça chave para abertura de dados públicos.

Agradecimentos

Esta pesquisa tem apoio financeiro da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (Processo #2017/02325-5) e CNPq.

Referências

- Attard, J., Orlandi, F., Scerri, S., and Auer, S. (2015). A systematic review of open government data initiatives. *Government Information Quarterly*, 32(4):399 – 418.
- Bertot, J. C., Jaeger, P. T., and Grimes, J. M. (2010). Using icts to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies. *Government Information Quarterly*, 27(3):264 – 271.
- Böhm, C., Freitag, M., Heise, A., Lehmann, C., Mascher, A., Naumann, F., Ercegovac, V., Hernandez, M., Haase, P., and Schmidt, M. (2012). Govwild: Integrating open government data for transparency. In *Proceedings of the 21st International Conference on World Wide Web, WWW '12 Companion*, pages 321–324, New York, NY, USA.
- Jaeger, P. T. and Bertot, J. C. (2010). Transparency and technological change: Ensuring equal and sustained public access to government information. *Government Information Quarterly*, 27(4):371 – 376. Special Issue: Open/Transparent Government.
- Kessler, C. and Kauppinen, T. (2015). Linked open data university of münster. In Springer, Berlin, H., editor, *Infrastructure and Applications*. University of München, Germany.
- Krumholz, H., Ross, J., Gross, C., and et al. (2013). A historic moment for open science: The yale university open data access project and medtronic. *Annals of Internal Medicine*, 158(12):910–911.
- Mutuku, L. N. and Colaco, J. (2012). Increasing kenyan open data consumption: A design thinking approach. In *Proceedings of the 6th International Conference on Theory and Practice of Electronic Governance, ICEGOV '12*, pages 18–21, New York, NY, USA.

Aplicando Transparência no Domínio de Informação Sobre Saúde na Internet: Uma Abordagem Visando o Cidadão

Nathália Miranda do Nascimento¹, Henrique Prado de Sá Sousa¹, Eduardo Kinder Almentero¹, Julio Cesar Sampaio do Prado Leite²

¹Departamento de Computação (DECOMP) – Universidade Federal Rural do Rio de Janeiro (UFRRJ) – Seropédica – RJ – Brasil

nathalia_miranda93@yahoo.com.br, {hsousa, almentero}@ufrrj.br

²Departamento de Informática – Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio) – Rio de Janeiro – RJ – Brasil

julio@inf.puc-rio.br

Abstract. *The internet has renewed the way communication and access to health information are made. However, the Internet has some difficulties related to the quality and access of health information. Transparency is a current demand of society, and the provision of more transparent health information by means of the internet demands special attention to its quality. The problem addressed in this work is related to the search and access to good quality health information available on the internet. To address this issue, an approach based on identification, selection and consolidation of information has been proposed. This approach culminated in the development of a prototype, which was used in a preliminary evaluation, the results of which are described.*

Resumo. *A internet renovou a forma como a comunicação e o acesso às informações de saúde são feitos. Porém, a internet está sujeita a dificuldades referentes à busca e qualidade da informação sobre saúde. A Transparência é uma demanda atual da sociedade, e o fornecimento de informações de saúde mais transparentes vai de encontro às tecnologias como a internet como um item de qualidade. O problema abordado neste trabalho está relacionado com a busca e com o acesso a informação de boa qualidade em saúde disponível na internet. Afim de tratar esta questão, uma abordagem baseada na identificação, seleção e consolidação de informações foi proposta. Tal abordagem culminou no desenvolvimento de um protótipo, que foi utilizado em uma avaliação preliminar, cujos resultados são descritos.*

1. Introdução

Pesquisas mostram que nos últimos anos o uso das TICs (Tecnologias de Informação e Comunicação) se intensificou e o processo de difusão da Internet se massificou, resultando no aumento da produção e propagação da informação sobre variados assuntos e em diferentes formatos. Entretanto, as informações disponibilizadas na internet não possuem garantia de qualidade e sequer de veracidade (MENDONÇA, 2013), não obstante, são frequentemente utilizadas para empregar falsas verdades (SIMÕES, 2016). Além disso, o grande volume de informações dificulta a localização de uma fonte confiável (TOMAÉL&VALENTIM, 2004).

Essa questão se torna especialmente relevante no contexto em que a internet passa a ser utilizada como uma ferramenta para busca de conhecimento sobre saúde, não raro, substituindo consultas médicas (COELHO *et al.*, 2013). Dados de 2012 indicam que 43% dos usuários de Internet buscaram informação relacionada à saúde ou aos serviços de saúde, representando a terceira maior atividade dos brasileiros na busca por informação online (BARBOSA, 2013), e pesquisas nacionais e internacionais indicam que o número de pessoas que utilizam a Internet para obter informações sobre doenças, tratamentos e até mesmo diagnósticos tende a aumentar cada vez mais (MORETTI *et al.*, 2012, AMANTE *et al.*, 2015). Logo, o acesso à informação de qualidade é parte integrante da promoção da saúde, que visa maior controle e poder de decisão ao usuário (GARBIN *et al.*, 2012).

No entanto, existe grande número de sites sobre temas de saúde-doença, bem-estar, qualidade de vida e promoção da saúde, produzidos por organizações ou indivíduos sem qualquer tipo de avaliação (PAOLUCCI, 2014). Muitos apresentam informações contraditórias, incorretas, incompreensíveis ou até fraudulentas. Há um consenso entre os autores sobre o risco apresentado pela busca de informação em saúde na Internet (MENDONÇA, 2013). Desta forma, a qualidade da informação tem se transformado em um novo desafio para a Saúde Pública no mundo (FIOCRUZ, 2016).

As buscas nas ferramentas tradicionais (ex. Google, Yahoo e Bing) não diferenciam as informações presentes nos sites em termos de qualidade, o que pode contribuir, por exemplo, ao acesso de informações dúbias, inconsistentes e sem garantia de veracidade. Entretanto, são as ferramentas mais utilizadas pela maioria das pessoas (independente de gênero e faixa etária), em contrapartida aos sites especializados, portais de notícia e fóruns (MINHAVIDA, 2016).

A partir das deficiências relacionadas a qualidade de informação na internet, mais especificamente no domínio da saúde, buscamos propor operacionalizações que facilitem o acesso a informações mais confiáveis e completas. No primeiro passo, avaliamos o conteúdo de diversos sites de saúde e identificamos as propostas existentes na área (Saúde) para reduzir as dificuldades de confiabilidade da informação. Este estudo culminou na definição de operacionalizações, as quais foram consolidadas em um processo que visa ampliar qualidade da informação de saúde na internet.

Tal processo foi baseado no uso de critérios para seleção de um conjunto de fontes de informação, o qual foi utilizado para a construção de um modelo conceitual consolidado, que subsidiou a criação de um protótipo de portal. Na construção do portal foram considerados alguns quesitos de Transparência (GER-PUC, 2018) para a operacionalização de elementos de interface. Posteriormente o portal foi avaliado por potenciais usuários, em comparação aos meios de busca tradicionais. Os resultados são resumidos neste artigo.

2. Processo Proposto

Os principais problemas identificados na busca por informações de saúde na internet estão relacionados à credibilidade das informações presentes em sites e aos diferentes focos dado ao tema que criam dispersão de informações sobre o assunto buscado. Além disso, alguns sites se aprofundam mais e outros menos em explicações sobre os determinados assuntos.

A escolha das fontes de informação que devem ser utilizadas em uma busca por dados sensíveis como os de saúde devem seguir critérios que mantenham algum grau mínimo de confiança acerca dos dados que ela disponibiliza. Um dos passos que realizamos foi a definição de critérios para a seleção das fontes de informação. Identificamos que existem certificados emitidos por órgãos de saúde a determinados sites que se encontram em conformidade com critérios específicos de qualidade da área (ex. (HON, 1999), (FIOCRUZ, 2018)). Outros critérios definidos foram a aceitação de informações provenientes de sites governamentais, instituições de saúde e universidades.

Partindo dos critérios, selecionamos algumas fontes de informação para aplicação neste estudo. Posteriormente mapeamos os conceitos oferecidos pelas fontes de informação. Esses modelos serviram de insumo para a construção de um modelo consolidado, que contém todas as informações disponibilizadas nas diferentes fontes. Na atividade de consolidação, foram avaliados os sinônimos para que não houvesse repetição de conceitos. O modelo consolidado foi utilizado como requisito para a definição de um protótipo com a capacidade de apresentar essas informações de forma integrada, permitindo ao usuário acessar o conteúdo das fontes de informação selecionadas através de uma busca facilitada. Posteriormente foi definido um protótipo que propõem uma forma de busca e apresentação dessas informações.

As atividades propostas criam um fluxo de identificação, seleção e apresentação de informações (Figura 1).

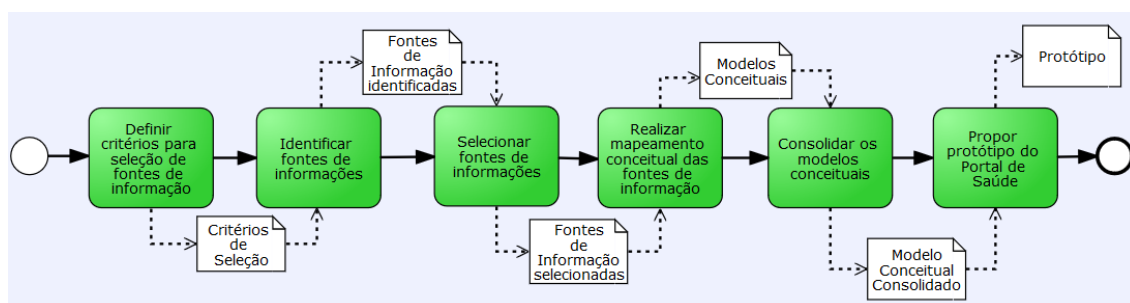


Figura 1 – Abordagem para ampliar qualidade da informação de saúde na internet

Para a definição do protótipo, algumas qualidades foram consideradas para apresentar as informações com maior grau de Transparência. Para isso foi utilizado o Catálogo de Transparência.

2.1. Identificação e seleção das fontes de informação

Para a identificação das fontes, foi realizada uma busca através do Google, mediante a busca por palavras-chave. Para simular o comportamento de um usuário comum, que busca informação na internet e desconhece a possibilidade do uso do vocabulário controlado e estratégias de buscas, optou-se por descritores de pesquisa escritos em português, sendo utilizados em várias combinações: portal de saúde; site de doença, site de saúde.

Na busca avançada os limites estabelecidos foram no modo “Pesquisa avançada”, opção “páginas em português” e país “Brasil”, e foram designados para busca avançada os mesmos descritores da busca simples em português mais a *string* de busca (“saúde”) AND (“site” OR “portal”) AND (“doença”), usando recurso de marcadores para

delimitar a pesquisa, tais como, as aspas e o sinal de positivo que servem como filtros na pesquisa avançada.

Após a identificação, foi obtida uma amostra de trinta e seis sites. A partir desta amostra, foi realizada a seleção de sites considerando os critérios de inclusão estabelecidos, que são:

Critérios de inclusão:

- Websites que tenham acreditação ou certificação da qualidade da informação na internet;
- Websites governamentais;
- Websites de instituições médicas;
- Websites que estejam entre os resultados apresentados até a página 3 do Google;
- Websites que contenham informações sobre doenças, por exemplo, sintomas e tratamento.

Entre os sites que foram retornados pelo sistema de busca e que não atendem aos critérios estabelecidos, temos como exemplo:

- Websites pessoais (blogs);
- Websites de teses e dissertações;
- Websites de ofertas de cursos referentes a temática;
- Website que abordam um único assunto;

Ao fim da seleção, os seguintes (Tabela 1) sites foram selecionados:

Tabela 1 – Lista dos Portais de Saúde Selecionados

Nome	Endereço
Portal sobre doenças raras e medicamentos órfãos - Orphanet	http://www.orpha.net/consor/cgi-bin/index.php?lng=PT
Minha Vida	http://www.minhavidacom.br/
DeCS - Descritores em Ciências da Saúde	http://decs.bvs.br/
Jornal Conceito Saúde	http://www.jornalconceitosauade.com.br/menu-doencas/
Ministério da Saúde – Portal da Saúde	http://portalsaude.saude.gov.br/
Criasaude	https://www.criasaude.com.br/

2.2. Mapeamento e consolidação das informações

A partir do conjunto de fontes de informação selecionado, foi feito o mapeamento dos conceitos presentes nos sites de saúde para reunir suas informações e consolidá-las, visando obter um modelo único do conhecimento presente nos sites. Com esse resultado espera-se ampliar o grau de completeza das informações.

A realização do mapeamento de informações consistiu em uma busca no site em seus diferentes níveis de páginas visando, em especial, identificar os conceitos chaves. Os recursos que mais forneceram informações foram os nomes de itens dos menus e os títulos internos das páginas. O mapeamento de cada site auxiliou a tornar o seu conteúdo mais transparente, o que nos permitiu analisar as diferenças nas informações presentes.

O conhecimento elicitado desses sites foi utilizado para construir modelos conceituais escritos na linguagem UML (*Unified Modeling Language*). O modelo conceitual gerado representa os conceitos referentes ao domínio do problema abordado. As diferenças de conceitos presentes nos sites (Figura 2 e Figura 3) explicitam a falta de padronização nas informações que são exibidas aos usuários.

Os modelos mostram que os sites abordam diferentes conceitos, e podem possuir mais ou menos detalhamento de informações sobre um mesmo conceito. Isso demonstra que o uso de uma única fonte de informação pode ser insuficiente para fornecer informações abrangentes sobre determinado assunto.

A partir dos modelos de cada portal foi construído um Modelo Conceitual Consolidado. A consolidação consistiu na união dos conceitos presentes nos sites, bem como a consolidação de termos que possuíam diferentes atributos. Basicamente conceitos e seus atributos foram somados ao modelo consolidado, eliminando-se repetições por palavras similares, especialmente entre os atributos. Este modelo não é apresentado aqui por motivos de espaço.

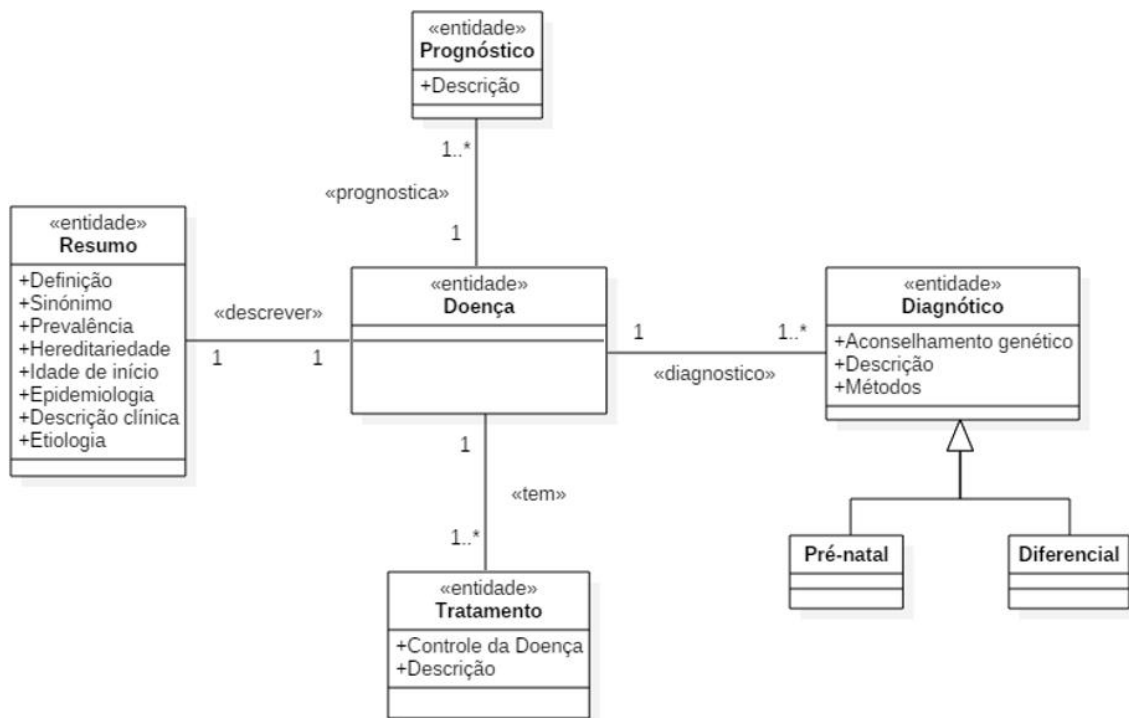


Figura 2 – Exemplo de modelo conceitual do Portal Orphanet

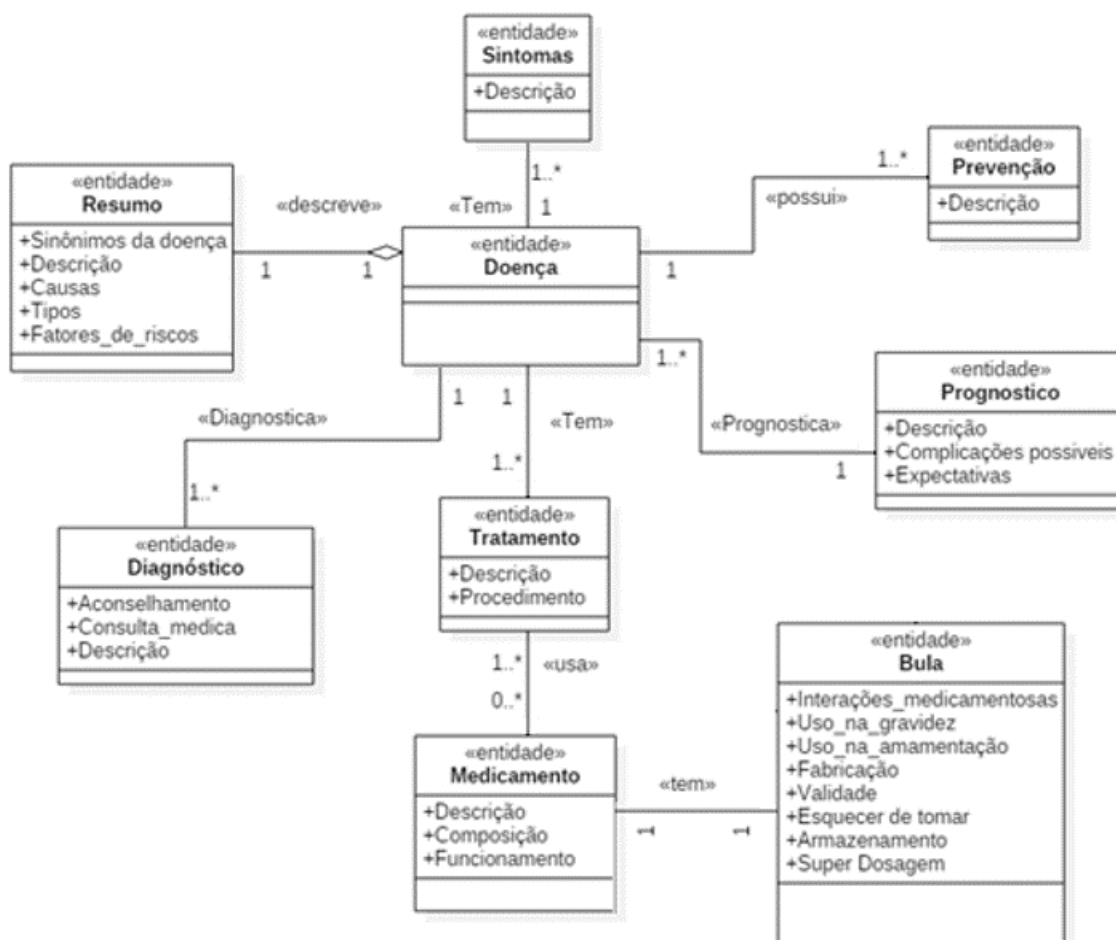


Figura 3 – Exemplo de modelo conceitual do Minha Vida

2.3. Construção do protótipo

A projeção do protótipo se baseou em operacionalizações que visam incrementar algumas das qualidades presentes no Catálogo de Transparência (GER-PUC).

O Catálogo de Transparência é fundamentado no GQM (*Goal, Question, Metrics*) (BASILI, 1992) e se baseia na decomposição de características de qualidade que, por sua vez, são refinadas em grupos de perguntas. As operacionalizações são formas de responder a estas perguntas e é daí que provêm a sua flexibilidade. Existem diversas formas de responder às perguntas existentes no catálogo e, conseqüentemente, inúmeras operacionalizações que contribuem para a transparência.

Neste trabalho foi feito a instanciação do catálogo de Transparência visando à transparência da informação sobre saúde ao cidadão. Mapeamos de forma básica algumas das contribuições positivas resultantes de atividades presentes na proposta de abordagem para ampliar qualidade da informação de saúde na internet, conforme ilustra a Figura 4. Por exemplo, a atividade “Definir critérios para fontes de informação” contribui tanto para Corretude quanto para Confiabilidade, ambas aplicadas à informação de saúde na internet.

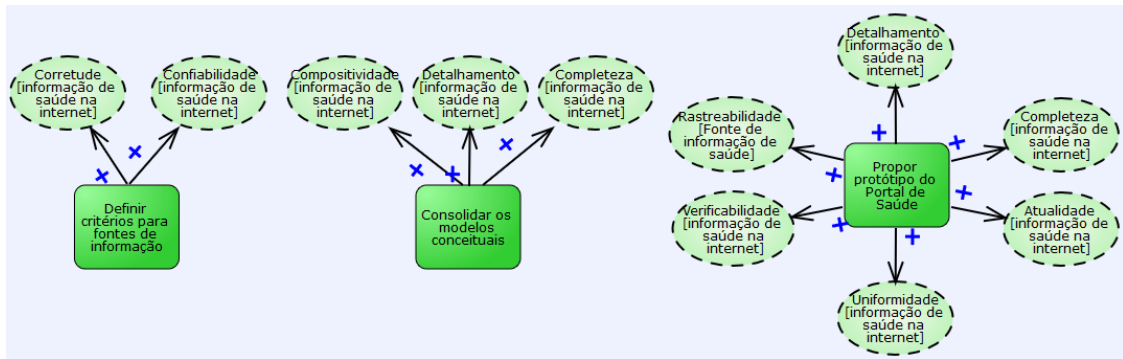


Figura 4 – Contribuições resultantes de atividades presentes na abordagem de transparência da informação de saúde na internet

No protótipo, em especial, foram operacionalizados outros atributos de transparência visando ampliar o grau de qualidade da informação. Por exemplo, para promover algum grau de rastreabilidade com as fontes das informações, foram incluídas as referências e elos para os sites originais, respaldando assim os conteúdos. Isso também ajuda a incrementar a qualidade de Verificabilidade da informação. Outra qualidade que se buscou incrementar foi a Atualidade, operacionalizada pela adição de informações como a data de publicação da informação e data da última atualização.

O grau de Completeza aumenta com a oferta de mais informações sobre diferentes aspectos de determinado assunto, possibilitado pela composição de informações que foi realizada na consolidação das fontes de informação (por consequência, também se alcança maior grau para a qualidade de Detalhamento). Também se contribuiu para a Uniformidade com a padronização na apresentação das informações. A Figura 5 apresenta o resultado parcial de uma busca no protótipo do portal de saúde, demonstrando as operacionalizações referenciadas anteriormente. Outras qualidades foram operacionalizadas no protótipo, mas não são apresentadas aqui por restrições de espaço.

Botulismo

Resumo da principais características da doenças

Resumo 1

Definição

- > Fonte: Descritores em Ciências da Saúde (DeCS)
- > Fonte: Minha Vida
- > FONTE: ORPHANET

O Botulismo é uma doença rara da junção neuromuscular, adquirida, caracterizada por paralisia flácida descendente causada por neurotoxinas botulínicas (BoNTs), inclui quatro formas com diferentes modos de aquisição.

Fonte: Orpha net

Por: Nome Editor
 Última Atualização: 11 Abril 2017
 Sobre: Tag 1, Tag 2, Tag3

Figura 5 – Apresentação parcial de um resultado de busca realizada no protótipo do portal de saúde

Nesta primeira versão do protótipo, foram implementadas a simulação de duas buscas por doenças (botulismo e artrose). Ainda não foram implementadas buscas automáticas nos sites mapeados. As buscas simuladas foram utilizadas na avaliação.

3. Avaliação

O protótipo foi avaliado por meio de uma pesquisa de opinião com potenciais usuários através de um questionário para verificar se o cidadão identifica as contribuições qualitativas que se buscou ao aplicar as operacionalizações propostas.

O questionário é composto de 25 questões fechadas de múltipla escolha e 1 questão aberta, para que os respondentes explicitem a sua opinião pessoal. Ao todo recebemos 21 questionários respondidos.

Para o preenchimento do questionário foi definido um procedimento de busca para que o usuário pudesse comparar a busca tradicional e a busca no protótipo do portal, e assim tivessem um conhecimento prévio suficiente para poder responder o questionário. Os procedimentos realizados pelos respondentes foram: (1) Fazer uma busca pelas doenças estipuladas (botulismo e artrose) através do Google, livremente, em seguida observar o resultado gerado, selecionar um ou mais sites para acessar e verificar as informações oferecidas por este(s) site(s); (2) Fazer uma busca no portal proposto pelas mesmas doenças estipuladas na etapa anterior. Observar o resultado gerado; (3) Responder ao questionário comparando as duas buscas e seus resultados.

Entre os principais resultados obtidos na comparação entre a busca tradicional e a busca no portal, apresentamos os seguintes resultados:

61,9% concordam que há maior facilidade na realização de buscas no portal.

66,6% concordam que encontraram informações mais completas no portal.

71,4% concordam que foi possível entender mais sobre o assunto a partir da busca no portal.

76,2% concordam que as informações apresentadas de diferentes fontes foram úteis para que os respondentes tirassem suas dúvidas.

76,2% dos respondentes concordam que tiveram maior confiança na pesquisa porque sabiam que o portal se baseia somente em sites certificados e de instituições governamentais.

Verificou-se que 90,5% dos respondentes que pesquisaram através do protótipo não utilizaram novas fontes de informação em sua busca. Já dentre os que utilizaram o Google, 47,5% foram a 2 ou mais sites para tirarem as suas dúvidas, indicando que o portal proporcionou maior eficiência e completeza de informações.

4. Conclusão

Este trabalho abordou os problemas relacionados com a busca e qualidade de informações de saúde disponíveis na internet. A fim de tratar este problema, foi proposto um processo que, em suma, proporciona a identificação, mapeamento, consolidação e apresentação de informações (através de um protótipo).

A construção do protótipo para apresentação das informações também considerou características de qualidade presentes no Catálogo de Transparência de Software (GER-

PUC). Além disto, foi possível verificar que características presentes no catálogo foram influenciadas pelo processo proposto neste trabalho. Por conseguinte, também foi proporcionada uma contribuição no âmbito do catálogo de transparência, uma vez que as ações aqui propostas podem ser incorporadas no catálogo como possíveis operacionalizações de requisitos não funcionais, permitindo seu reuso no futuro.

Também foi realizada uma avaliação com potenciais usuários através do protótipo. A avaliação permitiu, apesar de suas limitações, chegar à conclusão que há indícios de que o trabalho proposto contribui para melhoria da busca e qualidade de informações sobre saúde na internet.

Contudo, é importante destacar que o um dos resultados alcançados é um protótipo, ou seja, não houve de fato a implementação de um software, e sim uma simulação do comportamento. Isto pode ter influenciado a percepção das pessoas que o utilizaram durante o procedimento de avaliação realizado, o que, conseqüentemente, limita as nossas conclusões. Além disto, outro fator que limita as conclusões é o número de respostas do questionário recebidas.

Como trabalho futuro pretendemos desenvolver o portal, avaliá-lo novamente com um número maior de usuários, estimular a aplicação do modelo proposto, fazer avaliação do portal através do catálogo e usar o catálogo para avaliar outros sites. Além disso, intencionamos utilizar o processo proposto em outros domínios e, assim, avaliar a sua contribuição para um contexto de transparência de informação mais amplo.

Referências

- AMANTE, D. J., HOGAN, T. P., PAGOTO, S. L., ENGLISH, T. M., LAPANE, K. L., Access to care and use of the Internet to search for health information: results from the US National Health Interview Survey. *Journal of medical Internet Research*, v. 17, n. 4, 2015.
- BARBOSA, A. Pesquisa sobre o uso das tecnologias de informação e comunicação no Brasil: TIC Domicílios e TIC Empresas 2011. São Paulo: Comitê Gestor da Internet no Brasil, 2013.
- BASIL, V.R., Software modeling and measurement: the Goal/Question/Metric paradigm, 1992.
- COELHO, E.Q., COELHO, A.Q., CARDOSO, J.E.D., Informações médicas na internet afetam a relação médico-paciente?, *Revista Bioética*, 21(1), pp.142-149, 2013.
- FIOCRUZ – Fundação Oswaldo Cruz Inovação em Saúde para a sociedade. Selo de Qualidade da Informação em Saúde na Internet. 2017. Disponível em: <http://portfolioinovacao.fiocruz.br/?page_id=1450>. Acessado em: 20/03/2017.
- GARBIN, H.B.R., GUILAM, M.C.R., NETO, A.F.P., Internet na promoção da saúde: um instrumento para o desenvolvimento de habilidades pessoais e sociais. *Physis-Revista de Saúde Coletiva*, v. 22, n. 1, 2012.
- GER-PUC (Grupo de Engenharia de Requisitos da PUC-Rio), Catálogo Transparência, 2017. Disponível em: <http://transparencia.inf.puc-rio.br/wiki/index.php/Cat%C3%A1logo_Transpar%C3%Aancia>. Acessado em: 11 março. 2018.

- HON - Hon Foundation. Código de conduta para sites web em Medicina e Saúde (HONcode), 1999.
- MENDONÇA, A.P.B., Critérios de qualidade para sites de saúde: uma proposta, p. 97, Rio de Janeiro: ENSP/Fiocruz, 2013. Disponível em: <<http://www6.ensp.fiocruz.br/repositorio/resource/369214>>. Acesso em: 10 mai. 2016.
- MINHAVIDA, A Jornada digital do paciente, 2016.
- MORETTI, F. A., OLIVEIRA, V. E., SILVA, E.M.K., Acesso a informações de saúde na internet: uma questão de saúde pública?, Revista da Associação Médica Brasileira, v. 58, n. 6, p. 650-658, 2012.
- PAOLUCCI, R., A necessidade de uma revisão da avaliação da informação online, Mestrado em Informação e Comunicação em Saúde, IES:ICICT/FIOCRUZ. Congresso Brasileiro Interdisciplinar de Promoção da Saúde, 2014.
- SIMÕES, A.L., Hoax – o avanço tecnológico e a propagação de boatos, IX Simpósio Nacional ABCiber, PUC-SP, 2016.
- TOMAÉL M.I., VALENTIM, M.L.P. Avaliação de fontes de informação na internet. Londrina: Eduel; 2004.

Patrocinador Diamante



GOVERNO
DO RIO GRANDE DO NORTE

Patrocinadores Bronze



PREFEITURA DO
PARNAMIRIM
Cuidando de voos.

Apoio Financeiro



MINISTÉRIO DA
EDUCAÇÃO



nic.br
Núcleo de Informação e Coordenação do Ponto BR

cgi.br
Comitê Gestor de Internet no Brasil