

# Descobrimos Projetos de Lei relacionados a Transparência

Priscila Engiel, Roxana Lisette Quintanilla Portugal, Julio Cesar Sampaio do Prado Leite

Departamento de Informática – PUC-Rio  
R. Marquês de São Vicente, 225 – Gávea, 22451-900, Rio de Janeiro – RJ – Brasil  
{pengiel, rportugal, julio}@inf.puc-rio.br

**Abstract.** *Engiel et al., 2014 found the demand for transparency in VOTANAWEB site, which was done manually for projects in 2013. This article uses the same idea, however, proposes the use of text-mining techniques, thus covering 5011 bills published in the site. Thus, we identified 131 bills related to transparency, which were checked with regard to their endorsement by VOTANAWEB users.*

**Resumo.** *Engiel et al., 2014 verificou manualmente a demanda por transparência no site VOTANAWEB para os projetos de lei no ano de 2013. Esse artigo utiliza a mesma ideia, porém, propõe o uso de técnicas de mineração de textos possibilitando assim abranger os 5011 projetos de lei publicados no site. Como resultado, foram identificados 131 projetos de lei relacionados a transparência, e verificado a votação sobre a aprovação desses projetos de lei pelos usuários do site VOTANAWEB.*

## 1. Introdução

O processo legislativo inicia-se com a proposta de projetos de leis por membros do congresso, que após discussões na câmara legislativa, podem se tornar leis. O site VOTANAWEB disponibiliza o resumo de tais projetos de lei, e possibilita que a população avalie se concorda ou não, podendo comentar sobre esses. O trabalho de Engiel et al., (2014) analisou a demanda por transparência através de uma busca manual neste site, utilizando o catalogo de transparência [Cappelli, 2009] como referência para a busca.

As técnicas de recuperação de informação (IR), e extração de informação (IE) vem sendo utilizadas para extrair conhecimento em documentos com textos livres contendo linguagem natural (dados não estruturados) [Portugal, 2016]. Tais técnicas permitem agilizar a filtragem de documentos relevantes, e reduzir o tempo que levaria a leitura manual de uma pletora de documentos.

No momento da realização desta pesquisa, *VOTANAWEB*<sup>1</sup> possuía 5011 projetos de lei. Apesar de existir uma busca no site, essa não é precisa, pois as palavras de busca são transformadas na sua forma base, e.g. *controlabilidade* se torna *controle*. Nesse contexto, a palavra *controle* resulta em projetos de lei não diretamente relacionados ao conceito de transparência, o que não ocorreria se fosse utilizado exatamente a palavra

---

<sup>1</sup> <http://www.votenaweb.com.br/>

controlabilidade. Por isso, este artigo propõe o uso de técnicas de IR e IE para automatizar a procura de projetos no software *VOTANAWEB*.

## 2. Objetivos da Pesquisa

Nossa proposta é extrair de forma automatizada todos os projetos de lei relacionados a transparência, e verificar sua aceitação pelos usuários do site. Para isso, recuperamos todos os projetos de lei do site *VOTANAWEB* utilizando técnicas de (IR) e o catálogo de transparência [Cappelli, 2009].

Realizar uma ordenação dos projetos minerados baseado na aderência nas qualidades de transparência, e pela aceitação dos usuários no site. Isto para entender a demanda por transparência.

## 3. Contribuições Esperadas

O tipo de busca que é apoiado pelas qualidades do catálogo de transparência [Cappelli, 2009] baseia-se na geração de palavras dentro de um curto espectro de proximidade conceitual, ou seja, por exemplo, não procuramos apenas pela palavra *transparência*, se não por um conjunto de palavras que são relacionados com o conceito de transparência. Isto ajuda a ancorar leis cujo tópico é transparência, mas não necessariamente utilizam dita palavra.

Os projetos minerados foram ordenados segundo a aderência as qualidades de transparência. Cada qualidade do catalogo tem um peso, segundo o nível que está colocado no SIG (*Softgoal Interdependence Graph*) de Transparência (Figura 1). Se um projeto possui mais de uma qualidade os pesos são somados. Uma segunda ordenação é dada pela votação dos cidadãos de tais projetos em *VOTANAWEB*, mostrando a demanda por transparência.

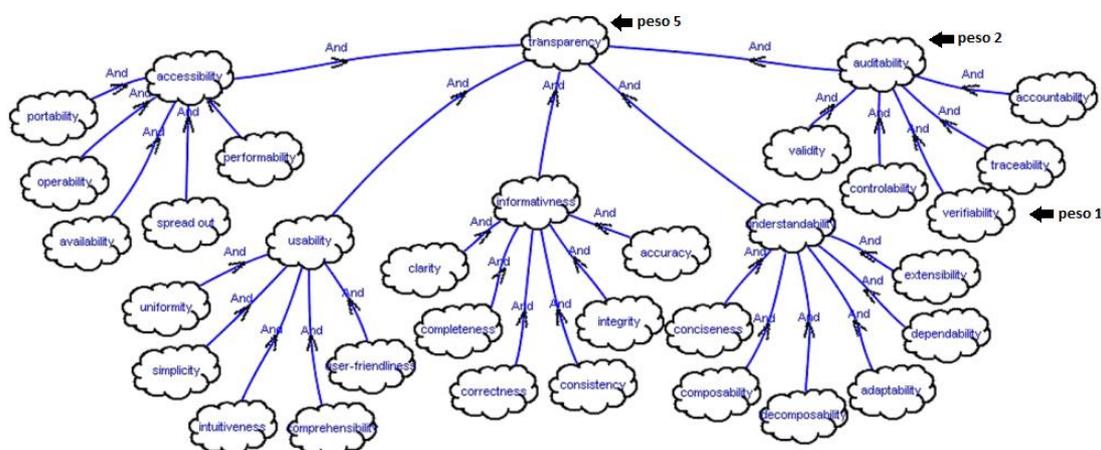


Figura 1. Rede de transparência (Leite e Cappelli 2008) - Versão 2, com pesos

#### 4. Resultados já alcançados

O SIG de transparência (Cappelli, 2009) descreve 33 qualidades, porém se procuramos apenas utilizando exatamente o nome das qualidades poderíamos não recuperar projetos importantes, pois é necessário considerar esses termos no contexto da lei. Por isso, os autores criaram uma lista de palavras-chave para cada qualidade do catálogo<sup>2</sup>, através de um *brainstorming*, e as combinaram para criar diferentes consultas. Por exemplo: Palavras chaves de *acessibilidade*: “acesso à informação”, “acesso”, “informação”, “dados”. As consultas utilizadas foram: “acesso à informação”, “acesso AND informação”, “acesso AND dados”. A lista completa de todas as palavra-chave e das consultas criadas foram disponibilizadas<sup>3</sup>.

A Tabela 2 apresenta o número de projetos retornado para cada qualidade. A característica que retornou mais projetos foi a de auditabilidade (211) seguida pela característica de acessibilidade (165). Os projetos filtrados também estão disponibilizados para análises futuras<sup>4</sup>. A qualidade com menos projetos foi *entendimento*, isso pode ter ocorrido pela dificuldade de traduzir em outras palavras certas qualidades, como por exemplo *concisão*.

Por fim, realizamos uma classificação dos projetos filtrados de acordo com o grau de transparência (quantidade de qualidades de transparência encontradas no projeto de lei), e pelo nível de aceitação dos usuários (proporção de sim (voto positivo)). Um projeto pode aparecer em mais de uma qualidade, quando isso ocorre, o projeto ganha mais peso, pois acreditamos que o projeto está mais comprometido com transparência. A lista dos 10 projetos mais transparentes é apresentada na tabela 1, juntamente com seu grau de aceitação que foi verificado manualmente no site *VOTANAWEB*. Podemos observar que 7 projetos têm mais de 80% de aceitação, apenas um projeto tem 18%, porém esse projeto fala sobre a exclusão da Controladoria Geral da União, sendo contrário a transparência, foi filtrado devido a uma negação.

Tabela 1. Top 10 – Projetos de Lei

Bill	Weight	Vote (yes)
0516-pls-79-2010.txt	10	95%
4940-plc-3497-2015.txt	10	94%
4029-pls-193-2013.txt	9	97%
2426-pls-387-2011.txt	9	92%
0499-pls-68-2010.txt	9	72%
0206-pls-82-2009.txt	8	93%
0164-plc-3305-2008.txt	8	86%
0786-plc-370-2010.txt	8	61%
0426-plv-8-2010.txt	8	18%
2963-plc-1080-2011.txt	7	96%

<sup>2</sup> <https://github.com/nitanilla/Mining-Brazilian-bills-related-to-Transparency/issues/11>

<sup>3</sup> <http://bit.ly/mining-br-bills>

<sup>4</sup> <https://github.com/nitanilla/Mining-Brazilian-bills-related-to-Transparency/issues/8>

Tabela 2. Projetos de lei classificados por qualidades de transparência

Qualidades de transparência	Número de Projetos de Lei	Total
Transparência	82	82
Usabilidade	14	102
Uniformidade	29	
Amigabilidade	8	
Simplicidade	24	
Operabilidade	10	
Intuitividade	7	
Adaptabilidade	7	
Desempenho	3	
Auditabilidade	97	
Validade	24	
Controlabilidade	62	
Verificabilidade	3	
Rastreabilidade	1	
Responsabilidade	24	
Accessibilidade	65	165
Portabilidade	11	
Disponibilidade	20	
Publicidade	69	
Entendimento	4	14
Dependência	2	
Compositividade	3	
Detalhamento	1	
Divisibilidade	4	
Concisão	0	
Informativo	17	67
Clareza	19	
Consistência	1	
Integridade	14	
Corretude	2	
Acurácia	10	
Atualidade	2	
Completeza	0	
Comparabilidade	2	

## 6. Conclusão

Entendemos que esse estudo continua o estudo iniciado por Engiel et al., (2014) e apresenta que apenas 2.6%, 131 projetos de lei, são relacionados a transparência. Porém a maioria desses projetos tem boa aceitação pelos usuários do site. É importante ressaltar que podem existir outros projetos relacionados a transparência que não foram mineradas com a estratégia proposta que utiliza o SIG [Cappelli, 2009] para compor consultas de busca. Com a mineração de textos, conseguimos analisar todo o universo dos projetos de lei presente no site, o que manualmente seria inviável. Ainda são necessários realizar ajustes nas palavras-chave para compor melhores consultas.

Acreditamos que existem diversas análises que podem ser realizadas sobre os projetos já filtrados, por exemplo identificar os tópicos frequentes nesses projetos de lei, ou o número de projetos relacionados a transparência por ano para verificar se existe alguma tendência. Acreditamos também que a mesma estratégia possa ser utilizada em outros

tipos de documentos para verificar a aderência à transparência, como por exemplo textos de leis já aprovados, em notícias e textos livres em linguagem natural em geral. Por isso, disponibilizamos todo o material em um ambiente transparente como o GitHub <http://bit.ly/mining-br-bills>.

É importante ressaltar que com essa estratégia automatizada se conseguiu identificar 13 de 28 projetos de lei que foram identificados de maneira manual [Engiel et al., 2014]. O ideal seria encontrar 100% dos projetos de lei. Por outro lado, essa estratégia identificou 7 novos projetos de lei para o ano 2013 que era o escopo do trabalho anterior.

## **Referências**

- Cappelli, C. “Uma Abordagem para Transparência em Processos Organizacionais Utilizando Aspectos”. Rio de Janeiro, 2009. 328 p. Tese de Doutorado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro. 2009.
- Leite, J.C.S.P.; Cappelli C. Software transparency. *Business & Information Systems Engineering*. 2010 Jun 1;2(3):127-39.
- Leite, J. C. S. P.; Cappelli, C. “Software Transparency”. *Business & Information Systems Engineering*, Springer, p. 127-139. 2010.
- Engiel, P., Leite, J.C.S.P, Cappelli, C. “Confirmando a Demanda por Transparência: Um Estudo Inicial sobre um Sistema de Avaliação de Projetos de Lei.” *Anais do WTRANS14 - Workshop de Transparência em Sistemas*, Londrina, 2014, Maio 28, 2014.
- Portugal, R.L.Q. *Mineração de Informação em Linguagem Natural para Apoiar a Elicitação de Requisitos*. Dissertação de Mestrado. Departamento de Informática, PUC-Rio. 2016.