

# RECUPERAÇÃO DA INFORMAÇÃO POR TÉCNICA WEBOMÉTRICA: Análise das menções web dos partidos políticos com representação no Senado Federal

*Information retrieval by webometric technique: Analysis of web mentions of the political parties with representation in Federal Senate*

**Eduardo Silveira<sup>1</sup>, Márcio Matias<sup>2</sup>**

(1) Doutorando no Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Santa Catarina, Florianópolis, edusilveira1985@gmail.com.

(2) Docente no Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Santa Catarina, Florianópolis, matias97@gmail.com.

## **Resumo:**

Apresenta a aplicação de técnica webométrica entre as conexões na web dos partidos políticos com representação no Senado Federal com o intuito de entender como se relaciona cada partido político na rede. Para tanto, a metodologia utilizada teve como abordagem o método quantitativo com a mensuração de cinco indicadores, no qual foram analisados o tamanho de site de cada partido político e a busca intercalada de dois partidos políticos de forma simples, por operador Booleano e por website com menção web. Entre os resultados verificou que todos os partidos políticos em estudo possuem conexões na web, que a busca com operador Booleano filtra mais páginas que uma busca simples. Em relação ao tamanho de site o PSDB é o que tem o maior número de páginas, assim como o que mais emite menções web dos demais partidos políticos. Ainda em relação aos resultados o coeficiente de Correlação Linear de Pearson apontou uma classificação muito forte para as variáveis Tamanho de site confrontada com Menções web recebidas e para Busca simples confrontada com Busca Booleana. Conclui-se que estudos entre recuperação da informação e webometria possuem correlação, assim como as conexões dos partidos políticos.

**Palavras-chave:** Recuperação da Informação; Webometria; Menção web; Correlação Linear; Partidos Políticos.

## **Abstract:**

It presents the application of webometric technique between the connections into the web of the political parties with representation in the Federal Senate with aims to understand how each political party relates on the network. To that end, the methodology used was the quantitative method with the measurement of five indicators, in which the size of each political party's site and the search of two political parties were analyzed in a simple way, by Boolean operator and by website with Web mention. Among the results it was verified that all the political parties in study have connections in the web, that the search with Boolean operator filters more pages than a simple search. In relation to the size of the site, the PSDB is the one with the largest number of pages, as well as the one that most emits web mentions of the other political parties. Still in relation to the results, the Pearson's coefficient Linear Correlation showed a very strong classification for the variables Size of the site confronted with received web Mentions and for simple Search confronted with Boolean Search. It is concluded that studies between information retrieval and webometric have correlation, as well as the connections of political parties

**Keywords:** Information Retrieval; Webometric; Web mentions; Linear Correlation; Political Parties.

## **1 Introdução**

A atual conjuntura política no Brasil fez com que milhares de brasileiros expressassem suas opiniões, seja nas ruas ou nas mídias sociais, de norte a sul do país. Nas ruas os indivíduos contra ou a favor das situações advindas da política nacional promoviam e participavam de manifestações, marchas e grito de ordem. Já nas mídias sociais, além de serem aliadas para

divulgação dos eventos oriundo do atual cenário, ocorreu em grande massa compartilhamentos de notícias, bem como a exteriorização de comentários e concepções individuais.

Muitos dos textos, notícias, vídeos compartilhados foram originados de páginas que não possuem credibilidade, fazendo muitas vezes a circulação de dados falsos e consequentemente promovendo o alcance de informações incorretas ao grande público.

Neste sentido, é primordial a procura e acesso a fontes de informação na web confiáveis. Para que cada indivíduo tenha acesso a bons recursos, se faz necessário uma busca de informações que resultam em páginas que tenham fidedignidade.

Essa fidedignidade pode ter bons resultados por meio de uma recuperação da informação elencadas em sistemas de recuperação da informação. Esses, tratam de

sistemas de operações interligadas para identificar dentro de um grande conjunto de informações (uma base de dados, por exemplo), aquelas que são de fato úteis, ou seja, que estão de acordo com a demanda expressa pelo usuário (ARAÚJO JÚNIOR, 2007, p. 72).

Uma base de dados que recupera vários tipos de fontes de informação é o buscador Google. Numa busca simples, dependendo do assunto em pesquisa, seja por uma palavra ou termo, o Google mostra ao usuário inúmeros resultados. E por varrer toda a web, apresenta resultados de páginas, mídias sociais, blogs, dentre outros suportes e que nem sempre recuperam 100% de informações confiáveis.

Uma forma de recuperar resultados mais autênticos se dá pela webometria, por meio de conexões entre websites e as menções web. De acordo com Orduña-Malea e Aguillo (2014) as menções web tem como características a descrição de um termo/palavra na web de forma textual ou hipertextual, sendo elas encontradas em qualquer parte do documento.

Estudos de menção web com a temática política foi abordado por Lin e Park (2011), que analisaram as menções de 18 membros da Assembleia Nacional da Coreia do Sul. Na coleta de dados foi identificada a visibilidade que cada membro apresentava na web, sendo as menções web localizadas em várias plataformas, como websites de notícias, blogs e imagens. A análise constatou que os membros que tem uma melhor carreira no mundo político apresentam melhor visibilidade no mundo virtual.

Os autores, em outro momento, investigaram a relação dos políticos coreanos do sul com o montante financeiro que receberam em doações do público. Na análise foi verificado que os políticos com mais

visibilidade na web tendem a receber mais doações financeiras, revelando assim uma correlação positiva entre a visibilidade das menções web e as doações recebidas (LIN; PARK, 2013).

As menções web podem ser analisadas em um âmbito geral (em toda a web) ou em nichos específicos, como no caso das duas situações analisadas por Lin e Park. Ao se deparar a um nicho ou grupo específico de websites ou um único website, as menções web podem trazer resultados mais confiáveis ao que se quer investigar.

Dessa maneira, a junção de uma menção web atribuída a um website junto ao buscador promove a combinação de artefatos que vem a culminar em resultados mais satisfatórios, pois será apresentado como resultado uma lista de documentos originados de um único website, que contém a menção web desejada. Assim, a fidedignidade das informações descritas nos documentos é certificada pela instituição que promove a divulgação do mesmo.

Em relação a política nacional, de dois em dois anos, os brasileiros acima dos 18 anos têm como obrigação participar das eleições municipais e nacionais, sendo a primeira na escolha de prefeitos e vereadores e a segunda na escolha de deputados estaduais, deputados federais, governadores, senadores e presidente.

Uma prática realizada durante a montagem de uma campanha política são as articulações de partidos para unir forças por meio de coligações partidárias. Essas coligações partidárias consistem “[...] na união de dois ou mais partidos que apresentam os seus candidatos em conjunto para uma determinada eleição.” (TODA POLÍTICA, 2015). Neste sentido, é muito comum que os partidos políticos mantenham relações entre eles no mundo real além das coligações partidárias, e essas relações podem também ocorrer no meio virtual por menções web.

## 2 Objetivos

O objetivo geral desta pesquisa é aplicar a webometria e identificar as conexões na web entre os partidos políticos com maior representação no Senado Federal, para tanto os objetivos específicos são:

(a) quantificar o tamanho do website de cada partido político;

(b) aplicar técnica de recuperação da informação por busca simples e por operador Booleano entre os partidos políticos;

(c) analisar as conexões entre os partidos políticos por meio de seus websites e suas menções web;

(d) calcular a correlação dos indicadores averiguados.

### 3 Procedimentos Metodológicos

A pesquisa pretende descrever e explorar por meio das técnicas de webometria os fenômenos entre as conexões dos partidos políticos pesquisados. Para tanto, o método de abordagem tem como característica o quantitativo, com a finalidade de quantificar e interpretar os dados resultantes da técnica apresentada.

A definição dos partidos políticos sucedeu em uma busca no site do Senado Federal (<http://www12.senado.leg.br>), em virtude de ser a entidade que compõe senadores com representação de todos os estados do Brasil. De acordo com a Senado (2017) o pleito é composto de 81 senadores com representação de 17 partidos políticos.

Para essa análise foram selecionados os cinco partidos políticos que tem a maior representação numérica no Senado, que totalizaram aproximadamente 65,43% (53 senadores) de todo o montante. A Tabela 1 apresenta os partidos selecionados.

**Tabela 1 – Partidos políticos selecionados**

Partido Político	Representação
Partido do Movimento Democrático do Brasil (PMDB)	21
Partido da Social Democracia Brasileira (PSDB)	11
Partido do Trabalhadores (PT)	9
Partido Progressista (PP)	6
Partido Socialista Brasileiro (PSB)	6
Demais partidos	27
Sem partido	1
<b>Totais</b>	<b>81</b>

Fonte: Adaptado do Senado Federal (2017).

Diante dos partidos políticos selecionados ocorreu a busca pelo website que corresponde a cada partido. Os respectivos websites estão representados no Quadro 1.

**Quadro 1 – Websites dos partidos políticos**

Partido	Website
PMDB	<a href="http://pmdb.org.br/">http://pmdb.org.br/</a>
PSDB	<a href="http://www.psdb.org.br/">http://www.psdb.org.br/</a>
PT	<a href="http://www.pt.org.br/">http://www.pt.org.br/</a>
PP	<a href="http://www.pp.org.br/">http://www.pp.org.br/</a>
PSB	<a href="http://www.psb40.org.br/">http://www.psb40.org.br/</a>

Fonte: Elaborado pelos autores (2017).

A partir das informações preliminares ocorreu o início da análise quantitativa. A coleta de dados compreendida nos três primeiros objetivos específicos teve como buscador escolhido o Google, realizada no dia 29 de agosto de 2017.

Para o primeiro objetivo específico 'tamanho do site' atribui a busca o conector 'site:'. O segundo objetivo específico direcionado à recuperação da informação foi utilizado a busca simples e o operador Booleano 'AND' seguidos do conector site:br; e para o terceiro objetivo específico que relaciona um website com a menção web aplicou o conector 'site:' seguido da menção web desejada.

O quarto objetivo específico teve como propósito a aplicação da correlação de Person sobre os indicadores coletados. Para esse objetivo foi utilizada a junção de duas variáveis e aplicadas no software SPSS versão 22.

Ademais, todos os dados da pesquisa foram tabulados em planilhas do Excel 2013 e quanto ao grafo apresentado em forma de rede utilizou os softwares Ucinet e Netdraw.

### 4 Resultados

Todos os dados foram recuperados pela sigla de cada partido político em virtude que os mesmos são mais conhecidos dessa maneira do que seus nomes por extenso. Os resultados apresentados estão divididos conforme os objetivos específicos propostos.

#### 4.1 Tamanho de site

O primeiro indicador analisado foi o tamanho do site, que consiste em averiguar o número de páginas que cada website possui. Para tanto, foi atribuído o conector 'site:' atribuído do website desejado conforme o exemplo: "site:partidopolitico.org.br". Os montantes estão apresentados na Tabela 2.

**Tabela 2 – Tamanho de site dos partidos políticos**

Estratégia atribuída	Tamanho de site
PMDB	42600
PSDB	162000
PT	18000
PP	493
PSB	9320

Fonte: Elaborado pelos autores (2017).

Os três partidos com maior representação política no senado são os três com a maior quantidade de páginas na web. Sendo o maior destaque para o PSDB que tem mais que o dobro de páginas somando os demais partidos analisados.

Também foi observado um fenômeno diferente no website do PP, cujo o número de páginas é bem inferior comparado com os demais partidos.

#### 4.2 Recuperação da informação Booleana

Com intuito de recuperar as conexões entre os partidos políticos em toda web foi submetida uma busca simples para cada par de partidos e uma busca com o operador Booleano AND aliado a expressão “site:br”, para recuperar as páginas em websites brasileiros.

O panorama das conexões da busca simples e com o operador Booleano podem ser apreciados na Tabela 3 e Tabela 4 respectivamente.

**Tabela 3 – Conexão pela busca simples**

Estratégia atribuída	Resultado
PMDB PSDB	3600000
PMDB PT	3160000
PMDB PP	1910000
PMDB PSB	2090000
PSDB PT	3120000
PSDB PP	1820000
PSDB PSB	5420000
PT PP	3490000
PT PSB	2090000
PP PSB	1400000

Fonte: Elaborado pelos autores (2017).

A seguir os resultados da estratégia de busca com o operado Booleano AND.

**Tabela 4 – Conexão pelo operador Booleano AND**

Estratégia atribuída	Resultado
PMDB AND PSDB	3440000
PMDB AND PT	3160000
PMDB AND PP	1920000
PMDB AND PSB	2140000
PSDB AND PT	3120000
PSDB AND PP	1820000
PSDB AND PSB	2140000
PT AND PP	3890000
PT AND PSB	1900000
PP AND PSB	1370000

Fonte: Elaborado pelos autores (2017).

O maior resultado obtido em conexão de duas siglas foi a dos partidos PSDB e PSB com 5420000 na busca simples e PT e PP com 3890000 na busca Booleana.

O segundo maior resultado das conexões foi dos partidos políticos que tem o maior tamanho de site (PMDB e PSDB), com 3600000 nas buscas simples e 3440000 nas buscas com o operador Booleano.

Das dez estratégias de busca sete obtiveram resultados diferentes, sendo quatro com maior expressão resultados de estratégias aplicados a buscas simples.

Baseado nos resultados apresentados entre as conexões atribuídas a busca simples e ao operador Booleano apresentamos a soma de páginas que cada sigla apareceu, conforme Tabela 5.

**Tabela 5 – Somatório individual de cada partido**

Partidos	Busca simples	Op. Booleano AND
PMDB	10760000	10660000
PSDB	13960000	10580000
PT	11860000	11860000
PP	8620000	9000000
PSB	11000000	7610000

Fonte: Elaborado pelos autores (2017).

Diante do resultado o que podemos perceber que quatro dos cinco partidos políticos tiveram na busca Booleana um montante de resultado menor, ou seja, por meio do operador Booleano ocorre uma redução nos resultados, que podem culminar a informações mais precisas e verídicas do que se procura, porém foram resultados muitos próximos.

Na busca simples o PSDB foi o maior somatório e na busca Booleana o PT é o partido com maior número de páginas. Já o

partido com o menor montante foi o PSB em ambas as buscas.

#### 4.3 Websites e Menção web

Com a finalidade de recuperar informações mais precisas acerca de conexões entre partidos políticos utilizou a técnica webométrica entre a junção do website e a menção web.

Para tanto realizou estratégias de busca com um dos websites dos partidos políticos com a sigla de outro partido político da população em estudo, conforme o exemplo: "site:partidopolitico.org.br Partido A". Desse modo os resultados apresentados estão dispostos em uma matriz conforme Tabela 6.

**Tabela 6 – Matriz das conexões por menção web**

Website	Menção web				
	PMDB	PSDB	PT	PP	PSB
PMDB		1200	2520	745	598
PSDB	9340		27000	3060	3560
PT	1740	3790		642	639
PP	10	8	5		3
PSB	567	412	803	133	

Fonte: Elaborado pelos autores (2017).

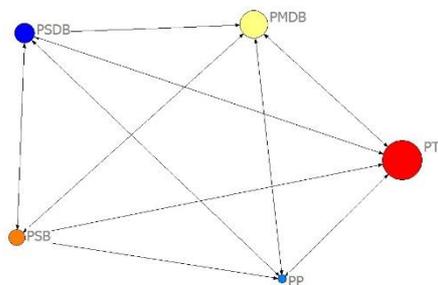
De imediato podemos afirmar que todos os websites possuem conexões por menção web como todos os partidos em estudo.

O PSDB foi o website que mais mencionou os demais partidos. Vale lembrar que o tamanho de site do PSDB colaborou com esse feito.

Já o PP por tem um tamanho de site inferior em relação aos demais foi o que menos mencionou os partidos estudados.

O panorama das menções web pode ser observado na Figura 1.

**Figura 1 – Conexões entre os partidos políticos**



Fonte: Elaborado pelos autores (2017).

A figura demonstrou que o PT foi o partido que mais recebeu menções, ou seja, é o partido com maior número de conexões junto de outro partido fora de seu website. De acordo com a Tabela 6 o PT é o mais mencionado em três websites (PMDB, PSDB e PSB).

O PP assim como é o website que menos menciona os partidos é também o que menos recebe menção em outros websites.

A Tabela 7 representa o quantitativo de todas as menções emitidas e recebidas pelas estratégias de busca atribuída de cada partido político.

**Tabela 7 – Quantitativo das menções web**

Partidos	Menções web	
	Emitidas	Recebidas
PMDB	5063	11657
PSDB	42960	5410
PT	6811	30328
PP	26	4580
PSB	1915	4800

Fonte: Elaborado pelos autores (2017).

Apenas o website do PSDB emitiu mais menções web dos demais partidos que recebeu menções deles. Sendo sua emissão mais que o triplo de todas as menções emitidas dos demais websites dos partidos.

Com 42960 menções emitidas significa dizer que aproximadamente 27% do montante de 162000 páginas do PSDB está relacionada com ao menos um partido da população em estudo.

Já o PP foi o menor website que mencionou outro partido político, com 26 menções web o montante representa aproximadamente 5% de relações com os demais partidos em sua página (total de páginas de 493).

#### 4.4 Correlação entre os indicadores

Após os indicadores coletados teve-se como proposta entender o nível de correlação de cada variável quantitativa. Para tanto será aplicado o teste de Correlação Linear de Pearson.

Segundo Barbetta (2014) a Correlação Linear de Pearson estará sempre entre o intervalo de 1 e -1. Quando mais próxima das extremidades é considerado uma correlação forte, quando próxima a zero é considerada

uma correlação fraca. Entre a correlação forte e fraca também existe a correlação moderada que são os carros em que o índice não se aproxima de 1, 0 e -1.

O Quadro 2 apresenta a diferenciação entre Correlação Linear bem fraca, fraca, moderada, forte e bem forte ou inexistência aplicada nesta pesquisa.

**Quadro 2 – Classificação de Correlação Linear**

Índice (r) (+ ou -)	Descrição
$(r) = 0$	Não existe correlação linear.
$0,001 \leq (r) < 0,199$	Correlação linear bem fraca.
$0,200 \leq (r) < 0,399$	Correlação linear fraca.
$0,400 \leq (r) < 0,699$	Correlação linear moderada.
$0,700 \leq (r) < 0,899$	Correlação linear forte.
$(r) \geq 0,900$	Correlação linear bem forte.

Fonte: Adaptado de SHIMAKURA (2006).

A aplicação de correlação linear teve como base cinco variáveis quantitativas, compostas pelos indicadores criados na análise de dados: Tamanho do site, Busca simples, Busca pelo operador Booleano AND, Menções Emitidas e Menções recebidas.

O Quadro 3 apresenta os cruzamentos das variáveis selecionadas.

**Quadro 3 – Classificação de Correlação Linear**

Variáveis	Código
Tamanho de site x Busca simples	CL1
Tamanho de site x Operador AND	CL2
Tamanho de site x Menções Emitidas	CL3
Tamanho de site x Menções Recebidas	CL4
Busca simples x Operador AND	CL5
Busca simples x Menções Emitidas	CL6
Busca simples x Menções Recebidas	CL7
Operador AND x Menções Emitidas	CL8
Operador AND x Menções Recebidas	CL9
Menções Emitidas x Menções Recebidas	CL10

Fonte: Elaborado pelos autores (2017).

Desse modo o Quadro 4 apresenta o resultado de cada variável confrontada e a classificação da correlação linear.

**Quadro 4 – Efeito da Correlação Linear de Pearson**

Código	Valor (r)	Classificação
CL1	0,833	Correlação Linear forte.
CL2	0,333	Correlação Linear fraca.
CL3	0,985	Correlação Linear bem forte.
CL4	-0,230	Correlação Linear fraca.
CL5	0,439	Correlação Linear moderada.
CL6	0,855	Correlação Linear forte.
CL7	0,174	Correlação Linear bem fraca.
CL8	0,335	Correlação Linear fraca.
CL9	0,794	Correlação Linear forte.
CL10	-0,175	Correlação Linear bem fraca.

Fonte: Elaborado pelos autores (2017).

Das 10 correlações averiguadas uma recebeu destaque com índice de correlação linear bem forte.

A CL3 que cruzou a variável Tamanho de Site e Menções web recebida ficou bem próximo de 1 com 0,985. Isto mostrou na análise que quando um website tem um número de páginas expressivo tende a emitir mais menções web.

O CL1, CL6 e o CL9 obtiveram como classificação os coeficientes de 0,833, 0,855 e 0,794 respectivamente, que corresponde a uma Correlação Linear Forte. Nesses três casos todas as variáveis estão presentes, demonstrando assim a forte ligação entre as variáveis estudadas.

Dois dos cruzamentos das variáveis apresentaram correlação negativa, vistos nos casos CL4 e CL10, a primeira como correlação linear fraca e a segunda com correlação linear bem fraca. Embora classificadas nesses índices, o negativo indica que se houver alterações nas variáveis a tendência é uma diminuir quando a outra aumentar.

## 5 Considerações Finais

Quanto ao objetivo principal da pesquisa que tinha como propósito aplicar a webometria nos partidos políticos com representação no Senado Federal foi atingido pois existe conexões na web entre eles.

Foi constatado que a maneira que se busca uma informação entre as conexões propostas ocorre resultados diferentes, sendo a técnica de webometria eficaz no que diz respeito a resultados mais precisos quando se analisa conexões entre duas entidades, pois a técnica confronta uma menção web e um website específico.

O que foi observado também que a webometria se correlaciona com a recuperação da informação, visto que nos confrontos das variáveis entre características de recuperação da informação com webometria os índices em metade dos cruzamentos foram expressivos, ou seja, entre correlação bem forte a moderada.

Em relação a temática Política, sugere-se estudos mais aprofundados de correlação dos partidos políticos com portais de notícias, bem

como informações do mundo real com o virtual.

### Referências

- ARAÚJO JÚNIOR, R. H. **Precisão no processo de busca e recuperação da informação**. Brasília: Thesaurus, 2007.
- BARBETTA, P. A. **Estatística**. Aplicações às Ciências Sociais. Florianópolis: Editora UFSC, 2014.
- LIM, Y. S.; PARK, H. W. How do congressional members appear on the web? Tracking the web visibility of South Korean politicians. **Government Information Quarterly**, v. 28, n. 4, p. 514-521, out. 2011.
- \_\_\_\_\_. The structural relationship between politicians' web visibility and political finance networks: a case study of South Korea's National Assembly members. **New Media & Society**, v. 15, n. 1, p. 93-108, fev. 2013.
- ORDUÑA-MALEA, E.; AGUILO, I. F. **Cibermetría**: Midiendo el espacio red. Barcelona: Editora UOC, 2014.
- TODA POLÍTICA. **Como funciona uma coligação partidária**. 2015. Disponível em: <<https://www.todapolitica.com/como-funciona-uma-coligacao-partidaria/>>. Acesso em: 09 jul. 2017.
- SENADO FEDERAL. **Senadores**. Disponível em: <<http://www25.senado.leg.br/web/senadores/em-exercicio>>. Acesso em: 11 jul. 2017.
- SHIMAKURA, S. E. **Interpretação do coeficiente de correlação**. 2006. Disponível em: <<http://leg.ufpr.br/~silvia/CE003/node74.html>>. Acesso em: 19 jul. 2017.