

# Descoberta de Conhecimento em Textos na política brasileira em 2014

Lennon Alves<sup>1</sup>, Felipe Gusmão<sup>1</sup>

<sup>1</sup>Departamento de Informática – Universidade Tecnológica Federal do Paraná (UTFPR)  
Av. Monteiro Lobato, s/n km04 – 84.016-210 – Paraná – PR – Brasil

LennonAlvesDias@gmail.com, FelipeGusmao@alunos.utfpr.edu.br

**Resumo.** *Este relatório apresenta a metodologia e resultados da atividade extra 2 da disciplina de Compiladores. Nele, será apresentado a implementação de um código-fonte para a ferramenta Flex com o intuito de buscar informações referentes às eleições brasileiras de 2014 e apresentação de gráficos que descrevem a situação dos candidatos do segundo turno com parâmetros (saúde, educação e segurança) utilizando bases de dados textuais de alguns sítios de notícia.*

## 1. Introdução

Analísadores léxicos que utilizam Flex lêem uma entrada padrão e podem ou não apresentar uma resposta. Aqui, será apresentado um arquivo em flex que lê um arquivo em linguagem natural em português e apresenta como saída, um pequeno relatório descritivo da quantidade de vezes que cada candidato do segundo turno das eleições presidenciais brasileiras de 2014 aparecem no texto, além de apresentar as incidências das palavras-chave: educação, saúde e segurança.

Foram coletados conteúdos textuais de páginas *web*, as quais: Uol Notícias: política, Veja Política, Yahoo e G1. Estes dados serviram como entrada do analisador e as saídas correspondentes foram base na geração de gráficos que mostram as incidências destas palavras-chave com os candidatos ao segundo turno das eleições brasileiras em 2014.

## 2. Metodologia

A seção 2.1 irá apresentar conteúdo pertinente ao arquivo flex o qual gera o código em linguagem C para depois ser compilado para um executável. Ele busca incidências dos nomes dos candidatos: Dilma e Aécio, além das palavras-chave: saúde, educação e segurança, se e somente se, houverem incidências dos candidatos no mesmo conjunto de dados de entrada.

### 2.1. Expressões regulares

Um arquivo flex possui três módulos principais e para a melhor compreensão, neste relatório, apresentar-se-á as expressões regulares que delimitam o escopo da busca. A seguir é apresentado o conjunto de expressões regulares onde as linhas de um a três representam as definições para a busca das palavras-chave. Analogamente, nas linhas 5 e 14 tem-se as definições que encontram os nomes dos candidatos à presidência.

Já nas linhas 6 a 8 e 15 a 17, encontram-se a relação do nome dos candidatos com as cada uma das palavras-chave. E nas linhas 10 a 12 e 19 a 21, definem-se relações diretas (i.e.: concatenações) de nomes dos candidatos com duplas de palavras-chave.

```

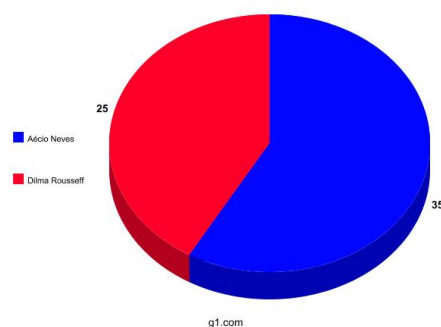
01| SAU sa(u|ú)de
   | EDU educa(c|ç)(a|ã)o
03| SEG seguran(c|ç)a
   |
05| BuscaA a(e|é)cio
   | SauA ({SAU}.*{BuscaA} | {BuscaA}.*{SAU})
07| EduA ({EDU}.*{BuscaA} | {BuscaA}.*{EDU})
   | SegA ({SEG}.*{BuscaA} | {BuscaA}.*{SEG})
09|
   | SauEduA ({SauA}{EduA} | {EduA}{SauA})
11| SauSegA ({SauA}{SegA} | {SegA}{SauA})
   | EduSegA ({EduA}{SegA} | {SegA}{EduA})
13|
   | BuscaD dilma
15| SauD ({SAU}.*{BuscaD} | {BuscaD}.*{SAU})
   | EduD ({EDU}.*{BuscaD} | {BuscaD}.*{EDU})
17| SegD ({SEG}.*{BuscaD} | {BuscaD}.*{SEG})
   |
19| SauEduD ({SauD}{EduD} | {EduD}{SauD})
   | SauSegD ({SauD}{SegD} | {SegD}{SauD})
21| EduSegD ({EduD}{SegD} | {SegD}{EduD})

```

O analisador resultante possui diversos contadores (variáveis numéricas) os quais se incrementam cada vez que a expressão definida é satisfeita.

### 3. Conclusão

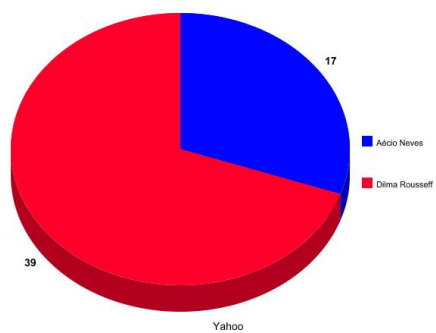
A partir dos gráficos das figuras 1, 2 e 3, têm-se os resultados da aplicação do analisador na base de dados de notícias G1, Yahoo e Veja política, respectivamente.



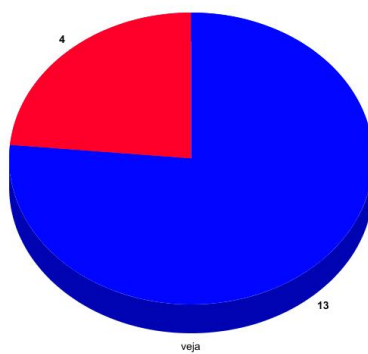
**Figure 1. Pesquisa G1**

Já na figura 4, tem-se a análise dos resultados da busca de todas as bases de dados utilizadas. E também, as incidências das palavras-chave utilizadas na busca.

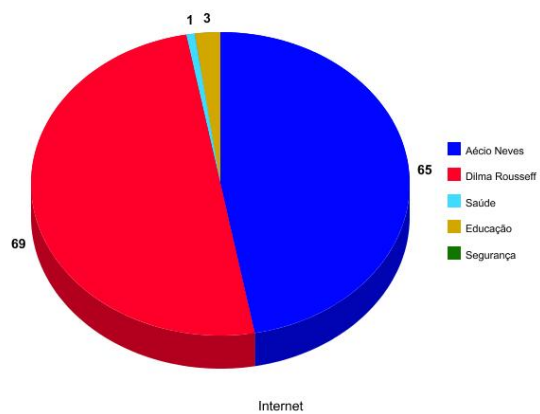
Através dos gráficos apresentados, pode-se perceber que a candidata Dilma possui mais incidências nos textos pesquisados que o candidato Aécio Neves. Também pode-se perceber que entre as palavras-chave, educação foi o termo mais discutido dentro dos textos.



**Figure 2. Pesquisa Yahoo**



**Figure 3. Pesquisa Veja política**



**Figure 4. Pesquisa: todas as bases de dados**