

## **TEMA**

(Tema Extraído de Maneira Automática)

Desenvolvedores

Fernando Antônio Asevedo de Nóbrega

Cleriston Aparecido Dantas

Thiago de Paulo Faleiros

## Índice

1- O que é TEMA? .....	pg. 3
2- Descrição do Processamento e Descoberta de Temas. ....	pg. 4
3- Manual do Usuário .....	pg. 5
3.1 Página Principal .....	pg. 5
3.2 Página do Tema .....	pg. 6
3.3 Página do Deputado .....	pg. 9

## 1- O que é TEMA?

**TEMA (Tema Extraído de Maneira Automática)** é um aplicativo web que apresenta todos os discursos proferidos pelos Deputados da Câmara de forma sucinta e organizada. Os discursos são apresentados por temas, sendo possível encontrar um tema específica, discriminar quais os temas discursado para cada Deputado, quais Deputados mais relacionados a cada tema, como esses temas são espalhados ao longo do tempo e quais as notícias relacionadas.

Os temas extraídos dos discursos são representados por conjunto de palavras com um valor semântico. Cada conjunto de palavras descreve um tema (contexto ou tópico) no qual um conjunto de discurso está abordando. Intuitivamente, considerando que um discurso é sobre um determinado tema, é esperado que palavras específicas possam aparecer com certa frequência: “rua”, “povo”, “movimento”, “população”, “sociedade”, “manifestações”, “publica”, “política”, “pais”, “brasileiro” – são palavras frequente em discursos sobre as manifestações populares que ocorreram em Julho de 2013, e juntas, essas palavras formam esse tema.

Os discursos são documentos textuais fornecidos pela Câmara dos Deputados dentro do projeto Dados Abertos. No total, foram coletados mais de cinquenta mil discursos. E desses discursos, foram identificados 200 temas distintos. O processo de descoberta de temas é um trabalho bastante custoso, sendo inviável sua realização por especialistas humanos devido a enorme quantidade de discursos. Dessa forma, é bastante útil a utilização de técnicas computacionais que realizam todo esse processo de forma automática.

O que existe de atual na literatura científica para a identificação automática de temas são os modelos probabilísticos. Esses modelos possuem forte embasamento matemático, em que, pela análise estatísticas das palavras, é possível descobrir qual tema descreve o conjunto de dados textuais. Específico ao sistema TEMA, foi utilizado o modelo LDA (Latent Dirichlet Allocation). O LDA foi criado por pesquisadores da Universidade de Stanford (David Blei, Andrew Ng e Michael Jordan) e possuem diversos trabalhos demonstrando sua eficácia na identificação automática de temas.

Foi utilizado o *Software* Abert MALLET (MAchine Learning for Language Toolkit – <http://mallet.cs.umass.edu/>) para a execução do algoritmo de inferência do modelo LDA. A ferramenta MALLET utiliza uma versão do algoritmo de amostragem de Gibbs para a inferência do modelo LDA.

## 2- Descrição do Processamento e Descoberta de Temas.

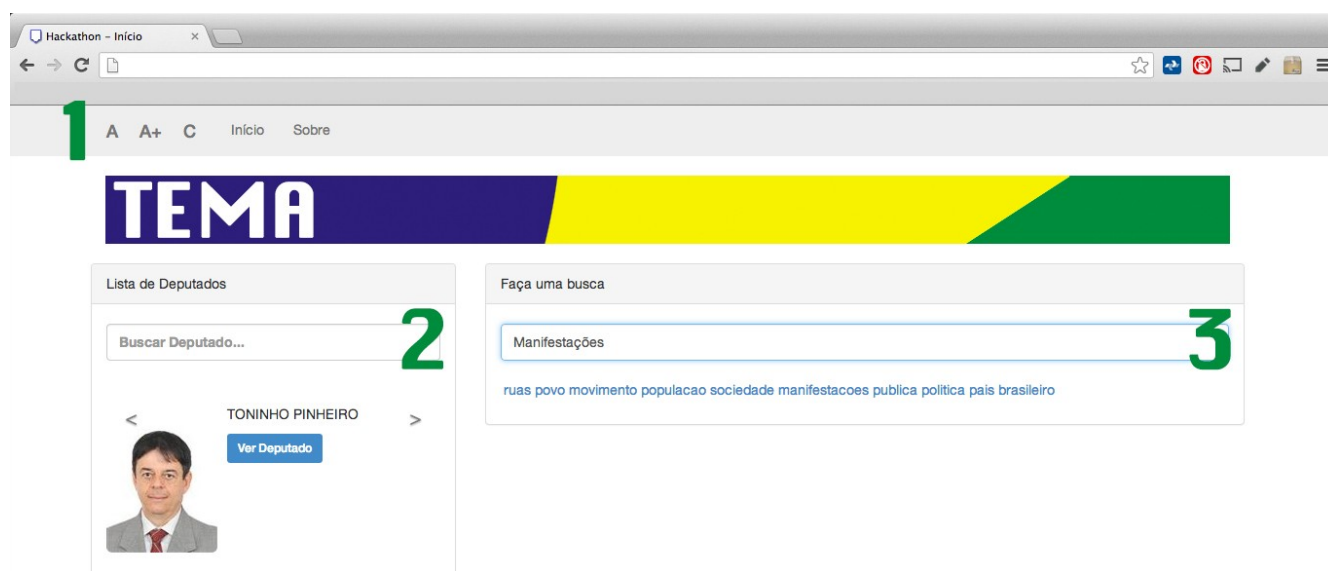
O processo de extração dos temas foi realizado da seguinte forma:

- *Coleta dos discursos (Oferecidos pela Câmara dos Deputados por meio da política de Dados Abertos – <http://www2.camara.leg.br/transparencia/dados-abertos>)*
- *Pré-processamento dos documentos textuais:*
  - *Limpeza de dados, remoção de caracteres especiais.*
  - *Remoção de stop-words.*
  - *Stemming*
- *Execução do algoritmo de inferência do modelo LDA com a base de discursos coletados (utilizando o toolkit MALLET)*
- *Transcrição das saídas do algoritmo de inferência para a base de dados.*

## 3- Manual do Usuário

O Objetivo desse Manual é apresentar as principais funcionalidades do aplicativo web TEMA para o usuário final. Nesse manual, serão apresentadas as principais telas do aplicativo com marcação numérica (números na cor verde) posicionadas próxima aos componentes de interação.

### 3.1- Página Principal



**Figura 1** – Página Inicial

**Figura 1** Item 1 — Para tornar a navegação agradável e mais acessível possível para todos os tipos de usuários, foram incluídos mecanismos de acessibilidade. Esses mecanismos estão em todas as telas do aplicativo. É possível ativar, por cliques do mouse ou acesso via tabulação, operações de aumento de fonte (A+), contraste (C) e

# TEMA

navegação alternativa (para a página inicial – Início).

**Figura 1 Item 2** – A lista de Deputados contém um campo de busca dinâmico. Para localizar um Deputado basta digitar o nome (nome completo ou nome parlamentar) e pressionar <Enter>. A cada letra digitada, a busca é efetuada e atualizada dinamicamente.

**Figura 1 Item 3** – É possível buscar um determinado inserindo a palavra relacionada. Ao digitar a palavra pressione <Enter> e todos os temas relacionados a essa palavra são apresentados na página.

## 3.2- Página do Tema



Figura 2 – Tela de Tema

**Figura 2 Item 4** – Uma vez escolhido um tema (busca do tema na tela Figura –

# TEMA

1 Item 3) são apresentados links para todos os tópicos. Clicando em algum link é apresentado a tela de temas (Figura 2). Na tela do tema é apresentado o gráfico descrevendo a ocorrência do tema (número de discursos sobre o tema) ao longo do tempo (mês em que o tema foi discursado) .

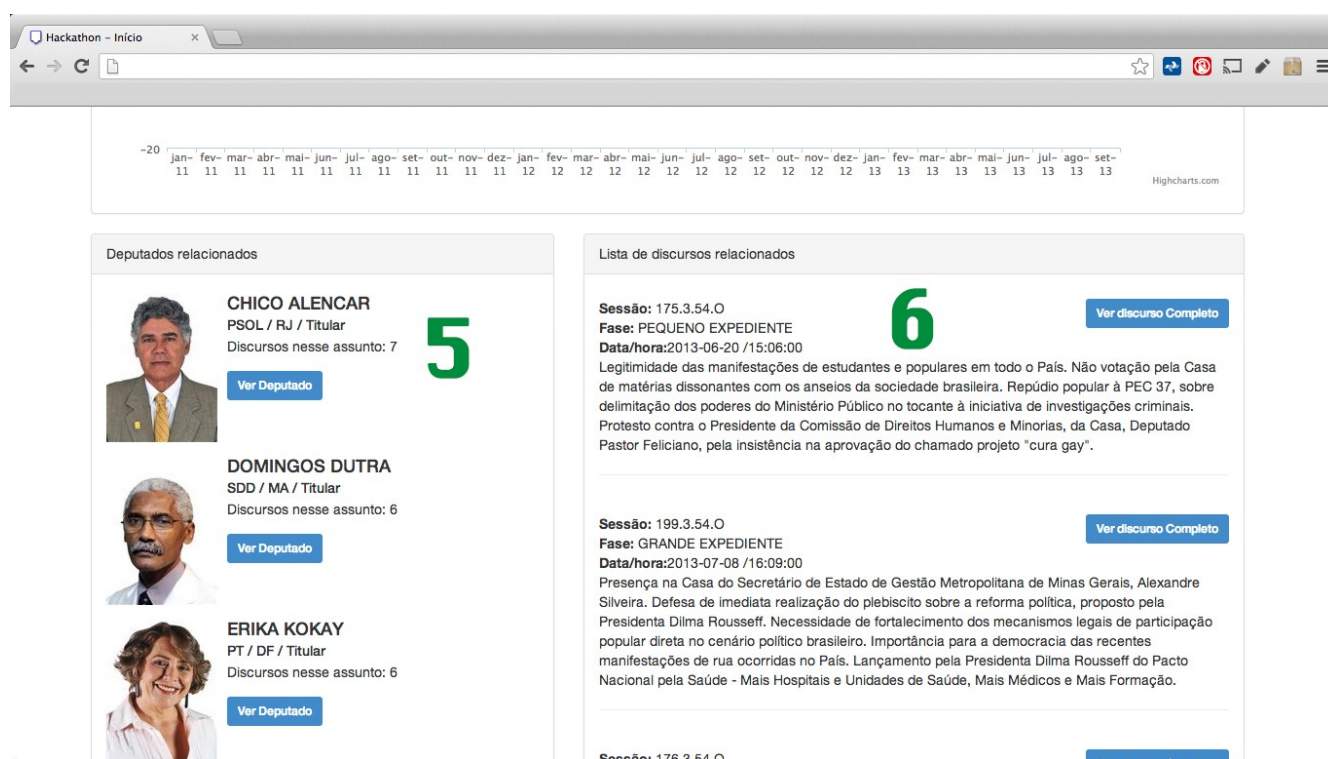


Figura 3 – Tela de Tema (Deputados e Discursos Relacionados)

**Figura 3 Item 5** — Na janela de Temas é possível ver todos os Deputados que discursaram sobre esse tema. Para cada Deputado, existe um botão “Ver Deputado” que irá para a tela de descrição do Deputado.

**Figura 3 Item 6** — São apresentados todos os discursos relativos ao tema. São listados os sumários dos discursos. O botão “Ver Discurso Completo” apresenta todo o discurso proferido.

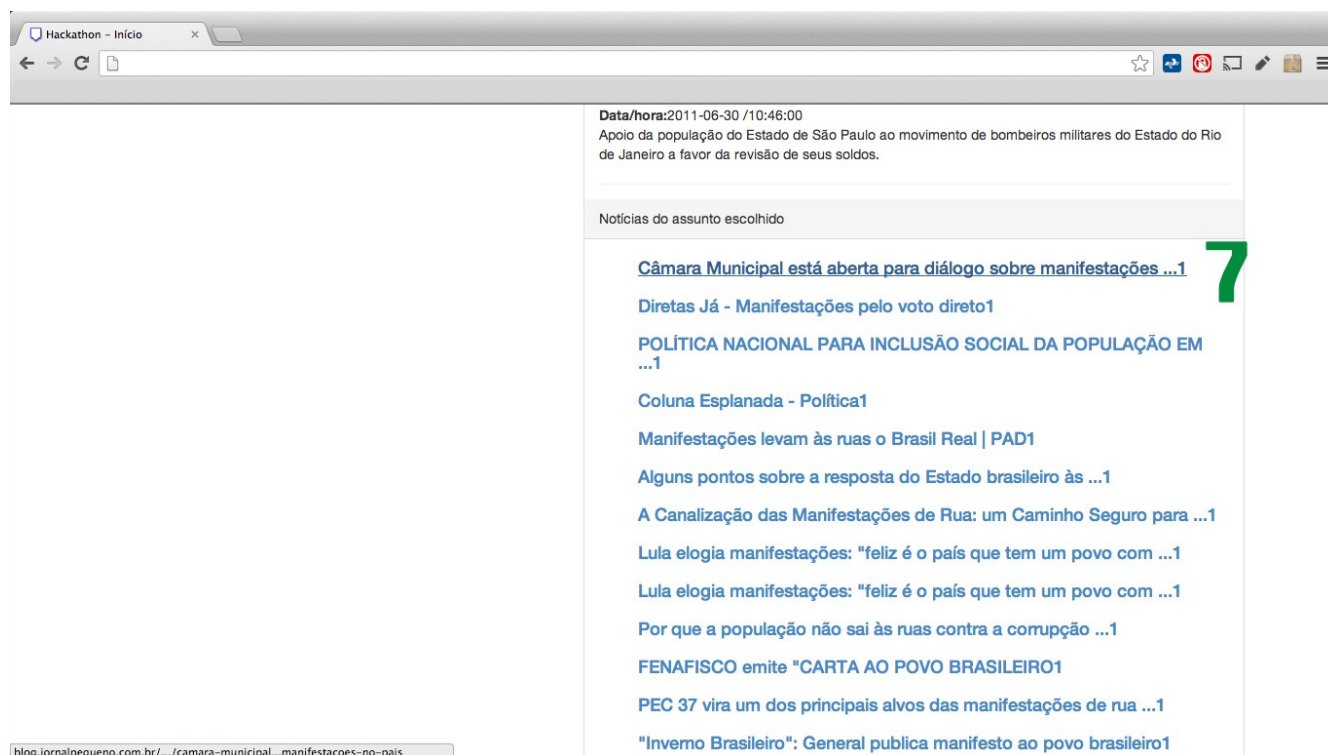


Figura 4 – Notícias Relacionadas ao Tema

**Figura 4** Item **7** — Aqui, na página de tema, são apresentadas as notícias relacionadas ao tema. Uma lista de links são apresentados e que podem servir como referência para um melhor entendimento do tema.



### 3.3- Página do Deputado

Deputado PAULO PIMENTA

**Nome civil:** PAULO ROBERTO SEVERO PIMENTA  
**Partido:** PT / RS / Titular  
**Gabiente:** 552, anexo 4  
**Fone:** 3215-5552  
**E-mail:** dep.paulopiment  
**Biografia (página externa)**

**8**

**Lista de Assuntos**

violencia morte vitimas assassinado familia pessoas  
 vida ocorre tragedia assassinato

[Analisar assunto](#) [Ver discursos](#)

tem ser pode respeito partido deve posicao ter  
 tenho dizer

**Lista de Discursos**

**Sessão:** 291.3.54.O  
**Fase:** ORDEM DO DIA  
**Data/hora:** 2013-09-25 /16:36:00  
 Solicitação à Presidência de inclusão na pauta de votações do projeto elaborado pela Casa em decorrência da tragédia ocorrida na Boate Kiss, no Município de Santa Maria, Estado do Rio Grande do Sul.

[Ver discurso Completo](#)

Figura 5 – Tela do Deputado

**Figura 4 Item 8** — Na tela do Deputado são descritas informações que relacionam o Deputado com todos os seus tópicos de discursados. O Item **8** contém a Tag-cloud com as palavras mais frequentes nos discursos do deputado.

Na Tela do Deputado (Figura 5) também são apresentados a Lista de Temas proferidos pelo Deputado e sua correspondente Lista de Discurso.

Na Lista de Temas, para cada tema específico, é possível, pelo link “Analisar Tema”, ver as informações deste tema (Figura 2), assim como é possível, pelo link “Ver discursos”, ver todos os discursos proferidos pelo Deputado relacionado a este específico tema.