

<p><b>Inteligência Analítica Aplicada ao Processo de Participação Social</b></p>
--

InWeb/UFMG ~~~ CEWEB/NIC.BR ~~~~Ministério da Justiça

**Arquitetura da Plataforma do Sistema de Inteligência Analítica**

Janeiro de 2016

## Sumário

1. Contextualização .....	3
2. Objetivos .....	5
3. Arquitetura da Solução .....	6
3.1. Arquitetura da Plataforma .....	6
3.2. Modelo de Operação e Funcionamento .....	7
3.2.1. Base de Dados .....	9
3.2.2. Wrappers (tarefas de enriquecimento de dados).....	12
3.2.2.1. Identificação de conteúdo relevante e de usuários influentes.....	12
3.2.2.2. Minerar padrões locais densos e aparar estes padrões para minimizar as sobreposições.....	13
3.2.2.3. Extração Automática de Tópicos.....	15
3.2.2.4. Análise de Sentimentos através de Aprendizado Supervisionado.....	16
3.2.2.5. Determinação de usuários influentes e conteúdos relevantes.....	17
3.2.3. Visualização de Dados .....	18
4. Sessão de usuário .....	20
Referências / Repositório do Projeto .....	21

## 1. Contextualização

O Anteprojeto de Lei de Proteção de Dados Pessoais é resultado de um amplo debate público promovido pelo Ministério da Justiça, em parceria com o Observatório Brasileiro de Políticas Digitais do Comitê Gestor da Internet no Brasil, que teve duração de cinco meses, recebendo mais de 14 mil visitas e obteve mais de 800 contribuições. Esses subsídios foram analisados e discutidos no âmbito do Poder Executivo entre os órgãos interessados e, posteriormente, consolidados em um texto final. Este texto foi submetido novamente a debate público a partir do dia 28 de janeiro de 2015, sendo finalizado no dia 5 de julho de 2015. Qualquer indivíduo, após se registrar na plataforma [www.participacao.mj.gov.br](http://www.participacao.mj.gov.br) pode contribuir comentando os artigos do anteprojeto de lei em discussão ou enviando contribuições consolidadas em PDF.

A proposta visa assegurar ao cidadão o controle e a titularidade sobre suas informações pessoais, de modo a concretizar o direito à inviolabilidade da intimidade e da vida privada, protegido constitucionalmente. O avanço da tecnologia da informação amplia enormemente o potencial de coleta, processamento e utilização de dados pessoais, o que representa, por um lado, uma oportunidade de geração de novos conhecimentos e serviços, mas, por outro lado, pode acarretar graves riscos aos direitos da personalidade do cidadão, ao acesso a serviços e bens, além de uma grande insegurança jurídica para o ambiente de negócios de tecnologia da informação existente no país.

A utilização cada vez mais intensa de dados pessoais na sociedade da informação cria um desequilíbrio entre os poderes dos indivíduos, titulares de seus próprios dados pessoais, e os dos utilizadores de tais dados, justamente pela quantidade de informações pessoais que as novas tecnologias são capazes de agregar e utilizar. Para que esses dados possam ser utilizados com fins transparentes e legítimos, ao mesmo tempo em que sejam garantidos os direitos de seus titulares, são necessárias normas e mecanismos institucionais que estabeleçam os parâmetros e limites deste manejo, até mesmo no momento de término dessa relação. Além disso, tendo em vista o caráter transnacional do fluxo dessas informações, cumpre indicar que qualquer discussão que verse sobre este tema abrange tanto as operações de tratamento de dados pessoais realizadas no Brasil, como aquelas realizadas no exterior, mas em que a coleta tenha ocorrido em território nacional.

A proposta abarca o tratamento de informações pessoais processadas tanto por meios automatizados como parcialmente automatizados, pelos setores público e

privado. Estão excluídos do âmbito de proteção da norma, no entanto, aqueles tratamentos de dados pessoais realizados para fins exclusivamente pessoais, bem como aqueles que tem por objeto o exercício regular da atividade jornalística e exclusivamente para tal fim, respeitando, ao mesmo tempo, os direitos individuais e a liberdade de imprensa. Quanto à regulação referente à segurança pública, esta deverá respeitar os princípios gerais estabelecidos no texto, porém contará com legislação específica posterior a esta proposta, além de ter a Autoridade competência para dispor sobre este tema.

Os direitos do titular, por sua vez, são explicitados, em particular com relação ao acesso, correção, dissociação e oposição ao tratamento de seus dados. Ademais, o anteprojeto estabelece normas específicas para o tratamento de dados cujo tratamento possa ensejar discriminação do titular (os chamados “dados sensíveis”, por se referirem a opção sexual, convicções religiosas, filosóficas ou morais, ou opiniões políticas, por exemplo), prevendo como regra geral que esses dados não devem ser tratados e que ninguém pode ser obrigado a fornecer informações de tal natureza a seu respeito, ressalvadas as hipóteses previstas em lei, assim como um regramento mais rígido quando o tratamento desses dados for permitido.

Assim, os dados pessoais merecem uma tutela forte e específica do ordenamento jurídico. O processamento dessas informações influencia diretamente a vida das pessoas, afetando oportunidades, escolhas e interações sociais, elementos que compõem o livre desenvolvimento da sua própria personalidade. Tendo isso em vista, é imperativo que haja um conjunto de princípios que norteiem o tratamento desses dados por terceiros, entre os quais podem ser destacados sua finalidade específica, adequada e necessária, além da regra de que o responsável pela coleta desses dados deva mantê-los em segurança, e que não os utilize para discriminação e permita o acesso facilitado ao titular.

Com o objetivo de dar efetividade à regulamentação sugerida, deverá ser objeto do debate a designação e a governança de um órgão como autoridade com competências quanto à proteção de dados pessoais no país. Tal órgão será responsável pela supervisão, auditoria, edição de normas complementares, aplicação de sanções administrativas, bem como pelo recebimento e encaminhamento de denúncias e reclamações dos cidadãos que sofreram danos em decorrência do tratamento indevido de dados pessoais e demais medidas necessárias para a implementação da legislação, sendo que sua organização reflete a complexidade e amplitude de sua tarefa.

A existência de um órgão para supervisão e controle do tratamento de dados se mostrou necessária na experiência de diversos outros países, tendo em vista o desafio de implementar uma legislação com um vasto âmbito de aplicação, bem

como pelo alto grau de especialização técnica necessária para lidar com as atividades relacionadas às tecnologias da informação.

Ainda, retomando o histórico que dá ensejo à proposta, pode-se citar quão necessária se demonstrou a criação de uma autoridade independente que tratasse da proteção de dados pessoais no caso da espionagem a autoridades brasileiras. No caso em questão, restou evidente que não havia qualquer arranjo institucional capaz de dar uma resposta de forma completa ao problema no Estado brasileiro.

Por fim, o texto abre espaço para que categorias profissionais e segmentos empresariais estabeleçam regras comuns, a título de boas práticas, outorgando ao mercado um grau necessário de autorregulamentação, sem prejuízo da observância aos princípios gerais da lei. A consolidação de um regime integrado de proteção de dados no Brasil mostra-se, assim, fundamental no ordenamento jurídico pátrio, de modo a possibilitar uma regulação integral do tema e a coesão de diversas iniciativas na área. Somente uma regulação geral assegurará a instituição de princípios harmônicos sobre o tema, proporcionará o controle dos riscos envolvidos no processamento de dados e garantirá a adequada reparação aos cidadãos pelos danos causados em decorrência do tratamento indevido de informações pessoais, garantindo a necessária segurança jurídica para a atividade empresarial e para a administração pública no tratamento de dados pessoais. Dada a importância não só do tema, mas a relevância da construção de metodologias de análises de participação social na construção legislativa que se faz cada vez mais presente na construção de diversas políticas públicas, há a necessidade de se pensar como rotular, validar e incorporar todas as informações de forma objetiva e analítica.

## 2. Objetivos

Este projeto tem por objetivo geral propor, implementar e avaliar uma plataforma de inteligência analítica para o processo de participação social, a qual contempla todas as etapas tradicionais do processo de descoberta do conhecimento em bancos de dados. Nesta direção, temos os seguintes objetivos específicos:

1. Coletar, extrair e armazenar dados de participação social;
2. Projetar, implementar e validar uma plataforma prototipal para carga e seleção de dados de participação social;
3. Projetar, implementar e validar modelos e algoritmos de inteligência analítica aplicados a dados de participação social;

4. Projetar, implementar e validar uma interface interativa e visual para inteligência analítica aplicada a dados de participação social.
5. Publicar como software livre a plataforma, incluindo modelos e algoritmos de inteligência analítica, para que possa ser utilizada e apropriada por outros grupos.
6. Realizar, juntamente com técnicos do MJ e do CEWEB, dois projetos piloto nos contextos das discussões em torno de Proteção a Dados Pessoais e Marco Civil da Internet.
7. Publicar relatórios detalhando o processo de construção e uso de cada modelo e algoritmo de inteligência analítica.

### **3. Arquitetura da Solução**

Nesta seção apresentamos a arquitetura da plataforma para acompanhamento de participação social através da internet, a qual é ilustrada na Figura 1 (Seção 3.1). Em seguida, na Seção 3.2, descrevemos o modelo de operação e funcionamento da plataforma, que é descrito através das tarefas que são executadas na plataforma.

#### **3.1. Arquitetura da Plataforma**

O ponto de partida da arquitetura é base de dados da consulta, obtida através de uma exportação do sítio que hospeda a mesma. Essa base de dados é então carregada em uma base de dados da plataforma ("Base de dados"), a qual suporta todas as atividades de inteligência analítica da plataforma. Essa base de dados deve suportar toda sorte de consultas de seleção, provendo as respostas a estas consultas no formato JSON e gerando "Dados selecionados". Os dados selecionados são a entrada das "Tarefas analíticas" a serem executadas sobre os dados.

A execução das tarefas pode gerar três tipos de informações: (i) indicar a necessidade de alteração nos parâmetros de seleção, com vistas a novo ciclo analítico; (ii) enriquecimento dos dados, os quais são incorporados à "Base de dados"; e (iii) metáforas visuais, as quais auxiliam o entendimento dos dados e do comportamento dos usuários no processo de participação social.

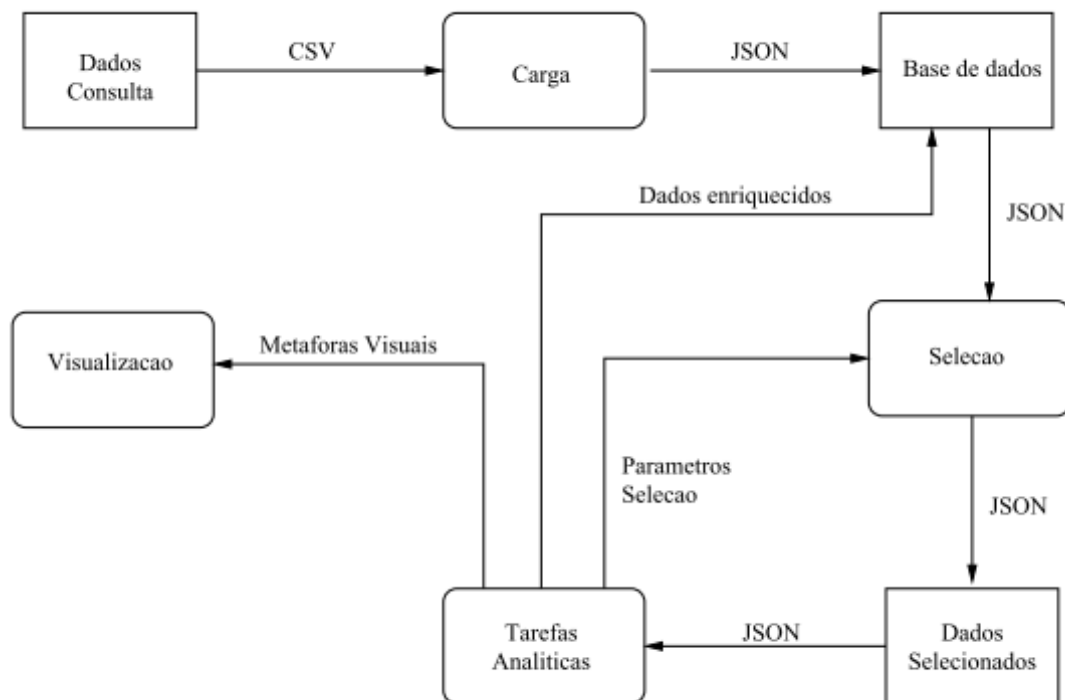


Figura 1 – Arquitetura da Plataforma

### 3.2. Modelo de Operação e Funcionamento

O Modelo de operação e funcionamento da plataforma desenvolvida é orientado a dados, pois assim fica flexível, expansível e escalável. Assim a arquitetura da plataforma, descrita na Seção 3.1, funciona por meio de tarefas que são executadas de forma independente, seguindo o modelo apresentado na Figura 2.

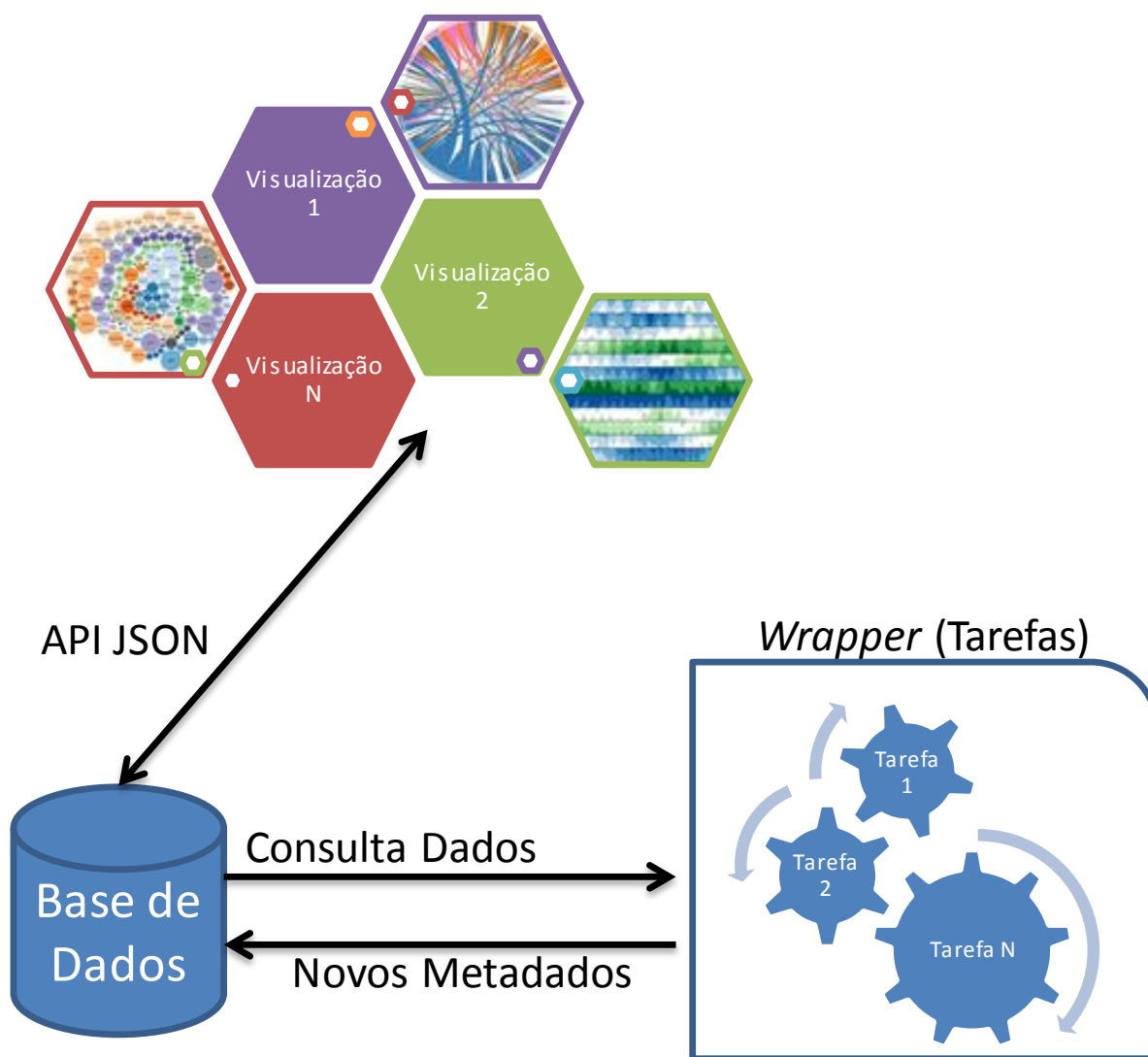


Figura 2 – Arquitetura Operacional - Modelo de Dados

Os elementos básicos desta arquitetura são:

- a) **Base de dados:** contém todos os dados da consulta pública, que são os seus itens comentáveis de um projeto de lei, os comentários (*posts* dos usuários) e os respectivos usuários participantes da consulta.
- b) **Wrapper de tarefas:** consiste de um módulo que empacota as tarefas de enriquecimento de dados, contendo uma entrada de dados, parâmetros de execução e saída (ou resultado do processo) que agrega valor aos dados, adicionando a informação de uma das seguintes formas:
  - Um novo campo ou coluna como metadado aos dados existentes;



- Um novo registro que adicione o metadado e relacione o mesmo a 2 ou mais entidades das existentes no modelo de dados (item comentável, usuário ou comentário).
- c) **Visualização de dados:** módulo que engloba as técnicas de visualização de dados que consomem direto da base de dados através de uma API JSON.

### 3.2.1. Base de Dados

Esta seção apresenta detalhes sobre a Base de Dados. Temos basicamente dois arquivos de dados.

O arquivo "dadospessoais-anteprojeto" contém os textos dos itens de lei presentes no anteprojeto, assim como a respectiva hierarquia e suas meta-informações. Os campos (colunas) são:

- commentable\_id: ID do item de lei.
- commentable\_parent: ID do item de lei acima na hierarquia, a qual o item pertence.
- commentable\_article: ID do item de lei do tipo artigo a qual o item pertence.
- commentable\_chapter: ID do item de lei do tipo capítulo a qual o item pertence.
- commentable\_axis: Número do eixo a qual o item pertence (1 a 13).
- commentable\_type: Tipo do item (artigo, alinea, inciso, paragrafo, seção, capítulo e geral).
- commentable\_name: Nome abreviado do item de lei.
- commentable\_text: Texto completo do item.

O arquivo "dadospessoais-comentarios-pdfs-filtrado" contém os comentários dos participantes dos debates, já filtrados e normalizados. Cada comentário está associado a algum item de lei da tabela "anteprojeto". Além disso, o comentário pode ter sido feito no website do debate ou através de um arquivo PDF. Os campos são:

- source: indica qual a mídia utilizada para enviar o comentário: "comment" indica um comentário feito no website do debate e "pdf" indica um comentário submetido via arquivo PDF.
- comment\_id: ID do comentário.
- author\_id: ID do participante que fez o comentário.
- author\_name: nome do participante que fez o comentário.
- comment\_parent: caso o comentário seja uma resposta direta a outro comentário, contém o ID do comentário respondido.
- commentable\_id: ID do item de lei que o comentário referencia.
- comment\_date: data e hora em que o comentário foi enviado.
- comment\_text: texto completo do comentário.

Além disso, os comentários têm campos adicionais que foram gerados através dos algoritmos de análise e processamento de dados:

- sentiment: valor calculado do algoritmo de análise de sentimentos;
- topic: tópico identificado pelo algoritmo de geração de tópicos.

### • **Dados de PDF**

Para processar os textos submetidos através de documentos PDF foi necessário realizar um processamento manual. O processo envolveu a segmentação dos textos contidos nos PDFs em porções de textos menores, de modo que cada porção de texto pudesse ser associada a um item de lei do anteprojeto. Em alguns casos, as porções de texto referenciam múltiplos itens de lei, portanto essas porções foram replicadas. Dessa forma, cada replicação referencia um único item de lei. A data e hora exatas dos comentários de PDF são desconhecidas e foram fixadas em "2015-08-01 10:00:00".

### • **Normalização e Filtragem dos Dados**

O banco de dados original dos comentários continha alguns comentários problemáticos (vazios, duplicados, etc). Esses comentários são prejudiciais para as análises, além de poderem gerar ruído nos resultados apresentados. Dessa

forma, foi realizada uma tarefa de filtragem e normalização dos dados. As tarefas realizadas foram:

- Limpeza do texto. Tarefas:
  - Conversão elementos de HTML (ex: "<" --> "<")
  - Remoção tags HTML (ex: "<br>" --> "")
  - Remoção de quebras de linhas e espaços em branco (ex: "\n" --> " ")
- Filtro de comentários. Foram removidos comentários:
  - vazios (em branco ou sem nenhuma palavra).
  - que continham apenas a palavra "teste".
  - duplicados (alguns usuarios fizeram, por erro dele ou do sistema, exatamente o mesmo comentário no mesmo item). Nesses casos apenas uma instância do comentário é mantida.
  - comentários de uma instituição (ABDTIC) foram replicados tanto no site quanto por PDF. No caso, mantivemos apenas os comentários do site.

Originalmente existiam 2015 comentários. Após a filtragem e normalização obtivemos 1856 comentários.

### 3.2.2. Wrappers (tarefas de enriquecimento de dados)

Esta seção descreve as atuais tarefas projetadas e desenvolvidas na versão atual da plataforma, que segue a arquitetura e modelo de dados propostos.

#### 3.2.2.1. Identificação de conteúdo relevante e de usuários influentes

##### a- Descrição

A tarefa de identificação de conteúdo relevante busca estabelecer quais itens do anteprojeto e quais comentários geraram maior reação de endosso/concordância e de reprovação/discordância entre os usuários. Complementarmente, a tarefa de identificação de usuários influentes pretende encontrar os usuários cujos comentários causaram tais reações.

##### b- Dados de entrada

Cada uma das tarefas é dividida em duas etapas: (i) preparação e (ii) execução do algoritmo ProfileRank.

A etapa (i) requer como entrada apenas a base de dados "dadospessoais-comentarios-pdfs-filtrado.json".

A etapa (ii) requer como entrada as saídas da etapa (i), que serão apresentadas mais abaixo.

##### c- Parametrização

Na etapa (i), o script "conteudo-relevante\_usuarios-influentes.py" não requer parametrização.

Na etapa (ii), o algoritmo ProfileRank solicita os seguintes parâmetros:

- diretórios de entrada "-i" e saída "-o";
- tipo de análise a ser realizada "-t", que pode assumir os valores "CONTENT" (para identificação de conteúdo relevante) ou "USER" (para identificação de usuários influentes);
- arquivo de entrada "--date", que pode assumir os valores "20000101" (para a identificação de conteúdo endossado), "20000202" (para a identificação de conteúdo reprovado) e "20000303" (para a identificação de usuários influentes);
- parâmetro fixo "-r", que deve sempre assumir o valor "0".

Exemplos:

- python ranking/pr.py -i dados/ -o dados/ -t CONTENT --date 20000101 -r 0 --> identificação de conteúdo endossado;
- python ranking/pr.py -i dados/ -o dados/ -t CONTENT --date 20000202 -r 0 --> identificação de conteúdo reprovado;
- python ranking/pr.py -i dados/ -o dados/ -t USER --date 20000303 -r 0 --> identificação de usuários influentes.

#### **d- Saída de dados**

A etapa (i) gera como saída três arquivos:

- 20000101, contendo os nomes dos usuários que endossaram conteúdo, os números de identificação dos comentários onde ocorreram os endossos e os números de identificação dos itens do anteprojeto ou dos comentários endossados;
- 20000202, contendo os nomes dos usuários que reprovaram conteúdo, os números de identificação dos comentários onde ocorreram as reprovações e os números de identificação dos itens do anteprojeto ou dos comentários reprovados;
- 20000303, contendo os nomes dos usuários cujos comentários receberam endosso ou reprovação e os números de identificação desses comentários.

A etapa (ii) gera como saída os seguintes escores:

- escore de endosso recebido, por conteúdo (item do anteprojeto ou comentário);
- escore de reprovação recebida, por conteúdo (item do anteprojeto ou comentário);
- escore de influência, por usuário.

Esses escores serão incorporados à base de dados na forma de novos campos na base de itens do anteprojeto, de comentários e de usuários.

### **3.2.2.2. Minerar padrões locais densos e aparar estes padrões para minimizar as sobreposições.**

#### **a- Descrição**

Dado qual cidadão comentou qual seção do anteprojeto de lei (artigo ou capítulo), se busca conjuntos de seções tal que a maioria destes cidadãos têm interesse na maioria destas seções (ou, dualmente, a maioria destas seções motivaram comentários da maioria destes cidadãos). Esperamos que esse trabalho possa guiar uma leitura mais racional dos comentários: em vez de ler juntos todos os comentários sobre um artigo (e perder a unidade de raciocínio de cada cidadão) ou juntos todos os comentários de um cidadão (e não poder contrastar as sugestões dele com

as dos outros cidadãos), os comentários podem ser lidos padrão por padrão. Como nada impede que alguns comentários sejam em vários padrão, "recortamos" os padrões para minimizar as sobreposições. Dessa forma, lendo os comentários padrão por padrão, não se lê muitas vezes os mesmos comentários.

#### **b- Dados de entrada**

Pares (cidadão, seção) cuja semântica é "o cidadão comentou a seção". O script `mine.sh` formata os dados brutos, como aqueles no arquivo `"dadospessoais-comentarios-pdfs-filtrado.csv"`, para abstrair todos os comentários abaixo dos artigos (por exemplo comentários sobre um alínea do artigo) como sendo comentários do próprio artigo.

#### **c- Parametrização (com um exemplo dos parâmetros usados)**

Atualmente, o script `mine.sh` (na pasta `"correlacoes"` do repositório) não aceita nenhum parâmetro. Ele é chamado com os dois arquivos que descrevem respectivamente as seções e os comentários:

```
$ ./mine.sh dadospessoais-anteprojeto.csv  
dadospessoais-comentarios-pdfs-filtrado.csv
```

Porém o tratamento de um maior conjunto de dados requeria um parâmetro para limitar os requisitos de tempo e de memória (ao custo de padrões de menor qualidade). Seria um valor numérico em argumento da opção `--ha` de `'multidupehack'`. Atualmente, esse valor não é usado pelo programa.

É também possível especificar um valor numérico em argumento da opção `--shift` de `'multidupehack'`. Quando menor o parâmetro, menos padrões maiores e menos densos são obtidos. Ao contrário, quando maior o parâmetro, mais padrões menores e mais densos são obtidos. A opção `--shift` não é usada no script `mine.sh` atual, o que significa `--shift 1`.

#### **d- Saída de dados**

Pares (conjunto de cidadãos, conjunto de seções) cuja semântica é "a maioria destes cidadãos têm interesse na maioria destas seções (ou, dualmente, a maioria destas seções motivaram comentários da maioria destes cidadãos)". Os padrões extraídos por `'multidupehack'` são recortados com `'d-peeler -m'` (ver o script `mine.sh`) para minimizar as

sobreposições.

[https://github.com/W3CBrasil/AI-](https://github.com/W3CBrasil/AI-Social/blob/master/correlacoes/biclusters.readable)

[Social/blob/master/correlacoes/biclusters.readable](https://github.com/W3CBrasil/AI-Social/blob/master/correlacoes/biclusters.readable) é um exemplo de saída legível. Ela foi obtida chamando o script `mine.sh` com os arquivos "dadospeessoais-anteprojeto.csv" e "dadospeessoais-comentarios-pdfs-filtrado.csv" em argumentos.

<https://github.com/W3CBrasil/AI-Social/blob/master/correlacoes/biclusters> é a mesma saída mas usando os identificadores dos cidadãos e das seções (o que facilita o seu pós-processamento, por exemplo para visualizar os padrões em uma página Web).

### 3.2.2.3. Extração Automática de Tópicos

#### a- Descrição:

Durante a discussão de um anteprojeto de lei, indivíduos, sociedade e organizações podem comentar e sugerir mudanças em cada um dos artigos e parágrafos do texto do anteprojeto. No entanto, um mesmo tópico de interesse pode ser discutido e comentado em diversas partes do texto. Nesse contexto, identificar comentários referentes a um mesmo assunto amplo é uma tarefa interessante, uma vez que pode oferecer uma visão mais global do processo de discussão e uma forma de sumarização dos tópicos mais discutidos. Nesse contexto, técnicas automáticas para extração de tópicos tem como objetivo agrupar informações de acordo com seu conteúdo semântico.

Em relação ao Anteprojeto de Lei de Proteção de Dados Pessoais, por exemplo, "consentimento de dados" e "proteção cibernética" podem ser considerados tópicos.

#### b- Dados de entrada

O técnica recebe como entrada os documentos dos quais se deseja extrair os tópicos.

Tanto os comentários quanto os documentos recebidos através do sítio de participação popular servem de entrada para o método.

#### c- Parametrização

O principal parâmetro recebido pelo método é o número de tópicos que devem ser extraídos.

O valor default utilizado foi 52, que corresponde ao número de artigos contidos no projeto mais a parte geral.

#### **d. Saída de dados**

Para cada comentário, a técnica gera um identificador, que corresponde a um tópico ao qual o documento pertence.

A partir desse campo, é possível visualizar os comentários referentes a um único tópico.

Além disso, cada tópico é representado por um conjunto de termos, que podem passar por uma avaliação qualitativa de um especialista para validação.

### **3.2.2.4. Análise de Sentimentos através de Aprendizado Supervisionado**

#### **a- Descrição**

Análise de sentimentos pode ser descrita como um conjunto de técnicas capaz de descrever a polaridade/opinião associada a um conteúdo textual. Especificamente, empregamos técnicas de aprendizado supervisionado para a tarefa de análise de sentimentos. Nesse sentido, o algoritmo de análise de sentimento produz um modelo de predição capaz de retornar a polaridade/opinião expressa em conteúdos textuais a partir dos padrões textuais nesse conteúdo. O modelo é produzido utilizando um recurso chamado dado de treinamento, o qual fornece as relações entre padrões textuais e polaridade/opinião. A saída do processo é a probabilidade de que o usuário que escreveu o conteúdo analisado tenha opinião positiva sobre o mesmo.

#### **b- Dados de entrada**

Dados de treinamento, composto por um conjunto de sentenças, parágrafos, artigos etc. Juntamente com cada conteúdo textual, há também a informação de classe, que indica a opinião de quem escreveu o conteúdo.

#### **c- Parametrização**

Não há parâmetros a serem variados.

#### **d- Saída de dados**

A saída de dados é composta de duas colunas por conteúdo textual (i.e., sentença, artigo, parágrafo etc.). A primeira coluna é um identificador, e a segunda coluna é a probabilidade de que o usuário que escreveu o conteúdo tenha opinião positiva sobre o mesmo.



### **3.2.2.5. Determinação de usuários influentes e conteúdos relevantes**

#### **a- Descrição**

Esta técnica tem por objetivo ordenar tanto usuários em relação à sua influência quanto os conteúdos em relação à sua relevância e impacto com base na disseminação desse conteúdo.

#### **b- Dados de entrada**

A entrada da técnica compreende as unidades de informação da consulta, as quais contêm duas informações de interesse: usuários e conteúdo. Os usuários são definidos unicamente pela plataforma e isso facilita o processo. O conteúdo é sempre gerado por um dado usuário, e pode ser encadeado com base na unidade de informação referenciada e na sua evolução cronológica.

#### **c- Parametrização**

A entrada consiste do arquivo de interações.

#### **d- Saída de dados**

A saída do processo ordena os usuários e seus conteúdos de acordo com a sua capacidade de formação de opinião e relevância, respectivamente. Essas ordenações podem também ser apresentadas por usuário, por exemplo um usuário de referência que se queira entender o contexto.

### 3.2.3. Visualização de Dados

Esta seção apresenta uma visão geral acerca do módulo de visualizações do sistema e das principais técnicas de visualização de dados implementadas. A **Figura 3** demonstra o processo de elaboração das visualizações, onde a partir de um arquivo de dados- no formato *json*, o código em *javascript*, carrega os dados, faz o processamento deles e apresenta a representação visual dos dados no navegador.

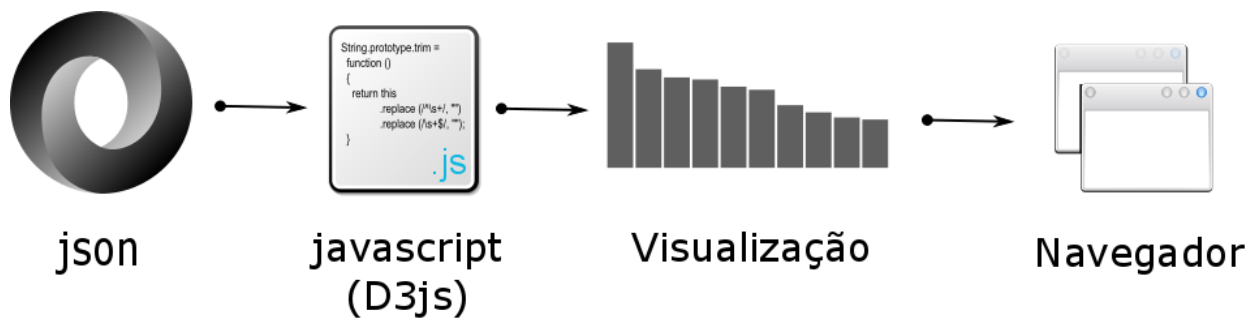
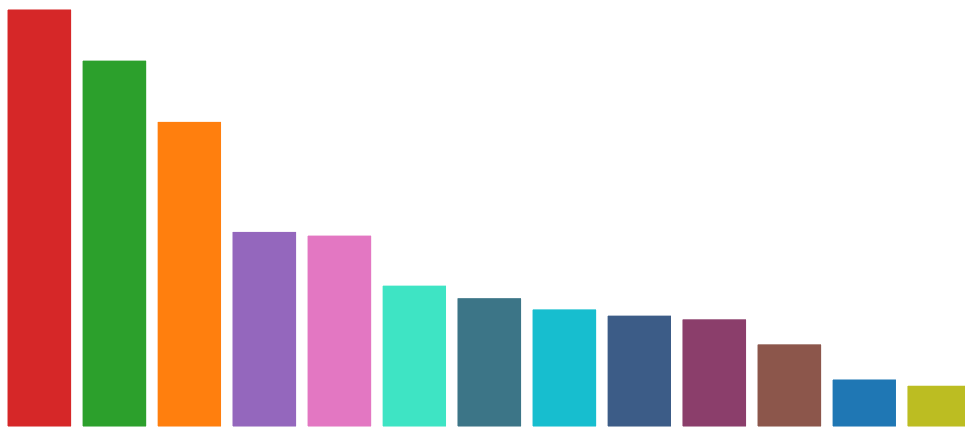


Figura 3: Processo de Visualização

#### Gráfico de Barras Verticais



O gráfico de barras é uma técnica clássica e bastante conhecida de visualização de dados que possui um conjunto de barras dispostas verticalmente. O eixo X do gráfico representa os elementos que estão sendo comparados e o

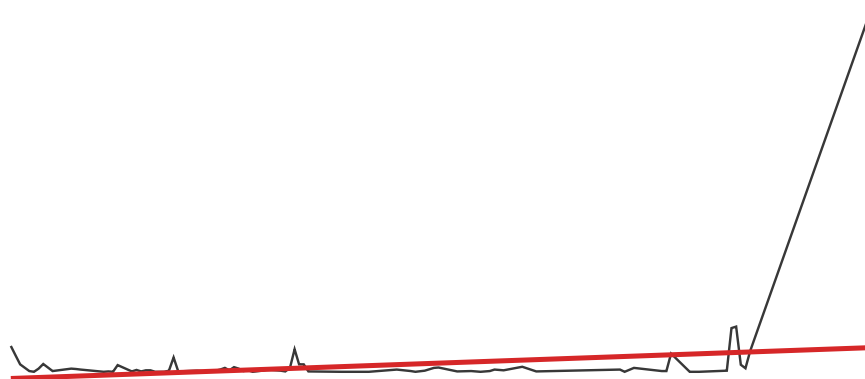
eixo Y os valores discretos. Através deste gráfico é possível fazer comparações e obter relacionamentos e padrões entre os dados.

### Gráfico de Barras Horizontais



Nesta variação do gráfico de barras as colunas estão dispostas horizontalmente, sendo assim, o eixo X do gráfico representa os valores discretos e o eixo Y os elementos que estão sendo comparados. Essa técnica é ideal para ver o relacionamento de *ranking* entre os itens que estão sendo comparados.

### Gráfico de Linhas



O gráfico de linhas é uma técnica de visualização de dados utilizada para exibir valores quantitativos ao longo de intervalos contínuos ou temporais. O eixo X geralmente possui valores categóricos ou escalas sequenciadas e o eixo Y os valores quantitativos. Através desta técnica é possível observar padrões e

comportamento dos dados ao longo do tempo, sendo frequentemente utilizado para exibir tendências e relacionamentos entre os dados quando agrupados com outras linhas.

### Grafo



Os grafos são técnicas de visualização de dados que possibilitam ver o relacionamento entre os dados. Visualmente temos os círculos representando as variáveis analisadas de um conjunto de dados e as linhas conectando cada um dos círculos e estabelecendo o relacionamento entre os dados.

## 4. Sessão de usuário

A execução de tarefas será controlada por um controle de sessão de usuários. Esse controle de sessão será realizado através de um controle de acesso com identificação (*login*) e senha individual.

Para o módulo de controle de sessão de usuário, pensamos em documentar as seguintes informações do usuário:

- Tarefas executadas (fluxos de execução);
- Histórico de atividades (ou ações) que realizou no portal;
- Histórico de novos dados (metadados que gerou com tarefas de enriquecimento de dados), gravando os dados para uso posterior em sua análise.

Para efeito de controle de acesso aos resultados gerados por um usuário e disponibilização desses novos dados enriquecidos por ele, definimos que isso será realizado através de níveis de permissão de acesso, que será inspirado na ideia de grupos de acesso que existe em sistemas do tipo Unix, que basicamente define os seguintes níveis de permissão aos resultados ou dados gerados com a ferramenta:

- Usuário: somente ele tem acesso;
- Grupo (ou contexto): usuário seta permissão de acesso aos dados disponibilizados a um grupo de usuários com quem ele interage;
- Público: fica disponível para todos os usuários da ferramenta.

### **Referências / Repositório do Projeto**

Repositório GitHub do projeto, a saber: <https://github.com/W3CBrazil/AI-Social/>

[1] Relatórios e explicativos das ferramentas utilizadas (Calaca, Kibana, etc.).

[2] Modelo de dados e arquivos da 1a consulta pública.

[3] Relatórios de aplicação prototipal das tarefas de inteligência computacional.

[4] Códigos e informações do portal v1 (dashboard) disponibilizado.

[5] Portal (dashboard) v1 de disponibilização dos dados e gráficos de caracterização dos dados da consulta pública, no seguinte endereço Web:

<http://homepages.dcc.ufmg.br/~diego.barros/research/prodape/index.html>