**PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL**

**ESCOLA POLITÉCNICA**

**CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

**FGApp - Sistema para acompanhamento de cargos comissionados no governo federal e filiação partidária.**

**JEAN CARLO RODRIGUES SCHUCHARDT BURDA**

**JOÃO BERTÉ**

Trabalho de conclusão I apresentado como requisito parcial à obtenção do grau de bacharel em Ciências da Computação na Pontifícia Universidade Católica do Rio Grande do Sul

Orientador: Prof. Avelino Francisco Zorzo

Porto Alegre, JUNHO de 2018

**FGApp - Sistema para acompanhamento de cargos comissionados no governo federal e filiação partidária.**

RESUMO

Encontramo-nos na era da informação. A cada segundo uma imensa quantidade de dados é gerada. Entretanto, muito desses dados são brutos e sem relação, logo se torna evidente que esses dados sejam analisados; a análise de dados possibilitará que informações sejam agrupadas, que novas relações entre os dados se tornem visíveis e com novas informações novas ações possam ser tomadas.

No Brasil, existem mais de 600 mil cargos em comissão, as informações desses dados são disponíveis, contudo, mesmo com o governo disponibilizando tais informações sobre esses cargos e suas remunerações respectivas, não há uma forma clara de observamos a movimentação que ocorre entre essas posições e outras eventuais relações que se tente encontrar. Sendo assim, o que dispomos atualmente são apenas dados crus e desconexos, tornando difícil encontrar alguma informação relevante. Utilizando de processamento e agrupamento de dados, é possível disponibilizar novas informações de forma mais visual e comparativa a fim de facilitar a compreensão sobre tais números, e verificar suas variações de forma conjunta com eventos e como foram afetados.

Este trabalho busca capturar dados e organizá-los utilizando conhecimentos do campo de ciência de dados. O novo conjunto de informações oriundo das análises e processamento de dados serão disponibilizados através de serviços *webs*, de forma acessível e transparente.

**Palavras-chave:** Processamento de Dados, Representação gráfica dos dados estatísticos, Mineração de Dados, Transparência Pública, Dados Públicos, Web Scraper.

**FGApp - Sistema para acompanhamento de cargos comissionados no governo federal e filiação partidária.**

ABSTRACT

We are in the information age. Every second an immense amount of data is generated. However, much of this data is crude and unrelated, so it becomes evident that this data could be analyzed; data analysis will enable information to be grouped, new relationships between data become visible, and new actions can be taken with new information.

In Brazil, there are more than 600 thousand positions in commission, data on these are available, however, even with the government making such information available about these positions and their respective remunerations, there is no clear way of observing the movement that occurs between these positions and any other relationships that may be encountered. So, what we have today is just raw and disconnected data, making it difficult to find any relevant information. By using data processing and grouping, new information can be made available in a more visual and comparative way to facilitate understanding of such numbers, and to check their variations in conjunction with events and how they have been affected.

This work seeks to capture data and organize them using knowledge of the field of data science. The new set of information from the analysis and processing of data will be made available through web services in an accessible and transparent way.

Keywords: Data processing, Graphical Re­presentation of Statistical Data, Data Mining, Public Transparency, Public data, Web Scraper.

LISTA DE TABELAS

[Tabela 1 Comparativo de recursos em trabalhos relacionados 17](#_Toc518244845)

[Tabela 2 Cronograma de Atividades 43](#_Toc518244846)

LISTA DE FIGURAS

[Figura 1 10](#_Toc518244817)

[Figura 2 Pedaço da tela principal do site Quem me Representa 11](#_Toc518244818)

[Figura 3 captura da tela do Twitter utilizado para divulgar as perguntas de suspeitas de gastos irregulares 13](#_Toc518244819)

[Figura 4 Mapa da Câmera 14](#_Toc518244820)

[Figura 5 Doações de Empresas para campanha de Deputados Fendereis 14](#_Toc518244821)

[Figura 6 Lista de Empresas Doadoras 14](#_Toc518244822)

[Figura 7 Top Doadores 15](#_Toc518244823)

[Figura 8 Quanto foi doado para cada partido? 15](#_Toc518244824)

[Figura 9 Qual o Destino das Doações para os Deputados Federais por Estado 16](#_Toc518244825)

[Figura 10 Distribuição de Funções Gratificadas no País 27](#_Toc518244826)

[Figura 11 Top Servidores que trocaram de Função 28](#_Toc518244827)

[Figura 12 Partidos políticos vs servidores 28](#_Toc518244828)

[Figura 13 Evolução do nº de servidores 29](#_Toc518244829)

[Figura 14 Site do TSE 30](#_Toc518244830)

[Figura 15 Consulta de Servidores 31](#_Toc518244831)

[Figura 16 Lista de funções e número de ocupantes 31](#_Toc518244832)

[Figura 17 Lista de servidores 32](#_Toc518244833)

[Figura 18 Detalhes sobre um servidor 32](#_Toc518244834)

[Figura 19 Remuneração de um servidor 33](#_Toc518244835)

[Figura 20 Dados em formato CSV 33](#_Toc518244836)

[Figura 21 Camada de extração 36](#_Toc518244837)

[Figura 22 Camada de pré-processamento 36](#_Toc518244838)

[Figura 23 Camada de análise de dados 37](#_Toc518244839)

[Figura 24 Diagrama de caso de uso – Usuário 38](#_Toc518244840)

[Figura 25 Diagrama de caso de uso – Administrador 39](#_Toc518244841)

[Figura 26 Diagrama de Atividade Camada de Extração 40](#_Toc518244842)

[Figura 27 Diagrama de Pré-Processamento e de Análise de Dados 41](#_Toc518244843)

[Figura 28 Diagrama de Atividade - Serviços e Usuário 42](#_Toc518244844)

LISTA DE SIGLAS

Dapp - Diretoria de Análise de Políticas Públicas

UX - User Experience

ISO - Organização Internacional para Padronização

FGV – Fundação Getúlio Vargas

TSE – Tribunal Superior Eleitoral

IBGE – Instituto Brasileiro de Geografia e Estatística

DAS – Direção e Assessoramento Superior

ONG – Organização não Governamental

KDD – Knowledge Discovery in Databases

URL - Uniform Resource Locator

CSV - comma-separated values

AWS - Amazon Web Services

UML - Unified Modeling Language ou Modelagem Unificada

SUMÁRIO

[1 INTRODUÇÃO 8](#_Toc518244788)

[2 TRABALHOS RELACIONADOS 10](#_Toc518244789)

[2.1 House of cunha 10](#_Toc518244790)

[2.2 Quem me representa 11](#_Toc518244791)

[2.3 Serenata de amor 12](#_Toc518244792)

[2.4 Câmera Transparente 13](#_Toc518244793)

[3 FUNDAMENTAÇÃO TEÓRICA 18](#_Toc518244794)

[3.1 CIÊNCIA DE DADOS 18](#_Toc518244795)

[3.1.1 A multidisciplinaridade de Ciência de dados 18](#_Toc518244796)

[3.1.2 Dimensões da Análise de Dados 20](#_Toc518244797)

[3.2 EXPERIÊNCIA DO USUÁRIO 22](#_Toc518244798)

[3.3 CONSIDERAÇÕES da seção 23](#_Toc518244799)

[4 PROPOSTA 24](#_Toc518244800)

[4.1 Necessidade 24](#_Toc518244801)

[4.2 Público-alvo 24](#_Toc518244802)

[4.3 Disponibilidade 25](#_Toc518244803)

[4.4 Legado 25](#_Toc518244804)

[4.5 Objetivos 25](#_Toc518244805)

[4.6 Ferramenta 25](#_Toc518244806)

[4.7 O risco do projeto 26](#_Toc518244807)

[5 Descrição do sistema 27](#_Toc518244808)

[5.1 Funcionalidades 27](#_Toc518244809)

[5.2 back-end do sistema 29](#_Toc518244810)

[6 Arquitetura do sistema 35](#_Toc518244811)

[6.1 Modelo de arquitetura em camadas 35](#_Toc518244812)

[6.2 Atores dos Sistema 38](#_Toc518244813)

[6.3 Diagrama de atividades 39](#_Toc518244814)

[7 CRONOGRAMA 43](#_Toc518244815)

[8 REFERÊNCIAS 45](#_Toc518244816)

# INTRODUÇÃO

O cenário político brasileiro tem gerado descontentamento de muitos brasileiros. Há cada dia a mídia apresenta mais e mais escândalos de corrupção e ações que desrespeitam o cidadão brasileiro; em consequência disso, o Brasil se encontra em um período de transformação.  Transparência e clareza, sobre os dados públicos, têm sido exigidas cada vez mais. O período em que o setor público se ocultava sobre a premissa da exacerbada burocracia e de dados públicos ocultos em imensas salas de arquivos estão em desvalia.

O volume de dados públicos disponíveis de forma digital tem aumentado. Sites como o Portal da Transparência, o Portal Brasileiro de Dados Públicos, o Tribunal Superior Eleitoral ou o Câmera Transparente disponibilizam muitos dados do setor público. Contudo, não é sempre que esses dados são apresentados de forma clara. Esses dados são disponibilizados em planilhas e páginas de difícil navegação, e utilizando muitos termos técnicos e próprios do ambiente de cargos públicos. Os dados também são disponibilizados de forma muito fragmentada, não é possível fazer agrupamentos por partido, estado, município dificultando uma visão mais ampla dos dados disponíveis.

Explorando esta dificuldade de grande parte da população em interpretar e analisar planilhas e dados de forma crua, este trabalho busca coletar e analisar dados e ao final exibir ao usuário uma nova percepção e novas informações sobre os conteúdos processados.

Apresentando os dados de maneira imparcial, como uma ferramenta de consulta de informações, para que o usuário consiga ver informações disponibilizadas através de sites do governo de uma maneira mais condensada e ampla. Ficando a cargo do usuário e de sua interpretação julgar as atitudes do Estado brasileiro.

Neste trabalho estamos retratando alguns trabalhos já existentes que abordam problemas relacionados, uma base teórica para o trabalho e as características de funcionalidade e de arquitetura de nossa aplicação. Na Seção 2 estão trabalho relacionados que foram caracterizados e retratados pelo grupo. A Seção 3 é dedicada para explicações dobre a base teórica que julgamos necessária para o desenvolvimento. A Seção 4 descreve a proposta de trabalho de conclusão de curso, relatando a necessidade de seu desenvolvimento, o público alvo, descrição breve da ferramenta, seus objetivos e riscos para o projeto. Uma descrição mais detalhada sobre a ferramenta, com suas funcionalidades e como irá coletar, tratar e expor os dados é escrita na Seção 5. Na Seção 6 é descrita a arquitetura do nosso trabalho. A Seção 7 apresenta um cronograma previsto das atividades para serem executadas no TC2.

# TRABALHOS RELACIONADOS

Alguns exemplos de trabalhos relacionados e que utilizam análise, mineração e visualização de dados referentes ao Estado e a governantes são os *sites* “House of Cunha” [1] e “Quem me Representa” [2], ambos do Laboratório *Analytics*, laboratório de pesquisa, desenvolvimento e treinamento na Universidade Federal de Campina Grande. O Dapp desenvolvido pela Fundação Getúlio Vargas e o Projeto *Open Source* idealizado por Irio Musskopf denominado Operação Serenata de Amor

## House of cunha

“House of Cunha” é um projeto realizado pelo Brasil de Fato que analisa a atuação de Eduardo Cunha na presidência da Câmara dos Deputados. Onde os votos do ex-presidente da Câmara dos Deputados, Eduardo Cunha, são comparados com os de outros candidatos e partidos. Os resultados das análises são disponibilizados em gráficos onde é possível visualizar a proximidade de candidatos e partidos em relação com os votos de Cunha Figura 1 apresenta um dos gráficos gerados na tela de apresentação deste *site*.

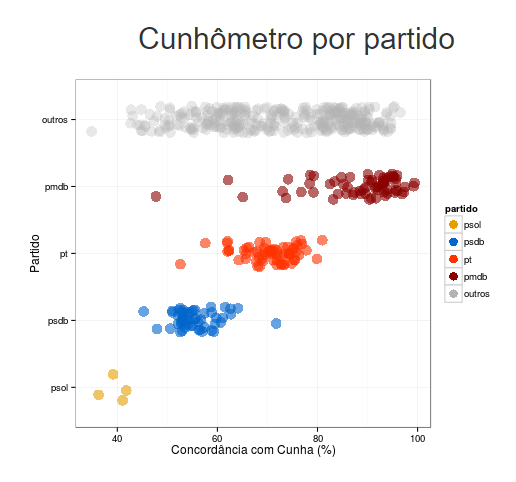


Figura 1

## Quem me representa

No projeto “Quem me Representa” você dá sua opinião em temas polêmicos e descobre quais deputados e partidos votaram concordando com suas opiniões. A opinião de entrada no *site* é comparada com os votos de todos candidatos eleitos de cada partido. Figura 2 apresenta a tela para compatibilidade entre suas escolhas e os votos dentro de cada partido.

A partir das escolhas, o algoritmo do site elenca os nomes dos deputados de acordo com o quanto estão alinhados com as opiniões do usuário, colocando ao lado de cada nome uma porcentagem de “compatibilidade”. É possível filtrar os resultados de acordo com os estados ou partidos dos congressistas [3].

Os dados do Quem Me Representa? foram extraídos do site da Câmara. Embora disponíveis a qualquer pessoa, estão espalhados pelas páginas das sessões e votações. A ideia da ferramenta é “simplificar” essas informações [3].

Marques, um dos criadores do site, quer que as pessoas discutam política “com base em dados e não achismos”. “Conheço gente que se surpreendeu com posições do seu deputado”, diz [3].



Figura 2 Pedaço da tela principal do site Quem me Representa

## Serenata de amor

Serenata de Amor é um projeto criado Pensando em fiscalizar gastos públicos da quota de atividade parlamentar, em busca de irregularidades. De acordo com os dados encontrados por eles, cada deputado gastou em média 266 mil Reais [4].

De acordo com Felipe B. Cabral, Operador na Operação Serenata de Amor, “O objetivo é criar um banco de dados que tem a função de oferecer respostas à diversas perguntas porque ele conecta muitas informações, e também aprende sozinho, observando a realidade. Pelo fato de aprender sozinho sobre corrupção, estamos chamando ele de Robô [4].”

O projeto consiste em uma inteligência artificial que verifica gastos suspeitos, ao encontrar algum gasto suspeito, um humano verifica se é ou não um gasto ilegal do parlamentar, conforme a resposta, a inteligência artificial vai aprendendo e melhorando sua assertividade na busca de irregularidades. O objetivo é que o “robô” tenha uma assertividade cada vez mais satisfatória e com menor dependência de respostas humanas.

A Serenata criou a Rosie: uma inteligência artificial capaz de analisar os gastos reembolsados pela Cota para Exercício da Atividade Parlamentar (CEAP), de deputados federais e senadores, feitos em exercício de sua função, identificando suspeitas e incentivando a população a questioná-los [5].

Esta inteligência, Rosie, utiliza o Twitter para realizar as perguntas, usuários cadastrados podem verificar e responder para ajudar a encontrar atividades suspeitas.



Figura 3 captura da tela do Twitter utilizado para divulgar as perguntas de suspeitas de gastos irregulares

Todo o projeto está público e qualquer órgão de combate à corrupção pode fazer uso dos algoritmos. Em qualquer lugar do mundo [4].

O objetivo a longo prazo é estar presente em todos os países democráticos e transparentes como uma ferramenta de controle social dos seus governantes [6].

## Câmera Transparente

Outro exemplo de projeto é Câmera Transparente desenvolvido pela o Diretoria de Análise de Políticas Públicas da Fundação Getúlio Vargas. A ferramenta apresenta um mapa da Câmera demonstrando as bancadas e quem ocupa cada cadeira (Figura 4).

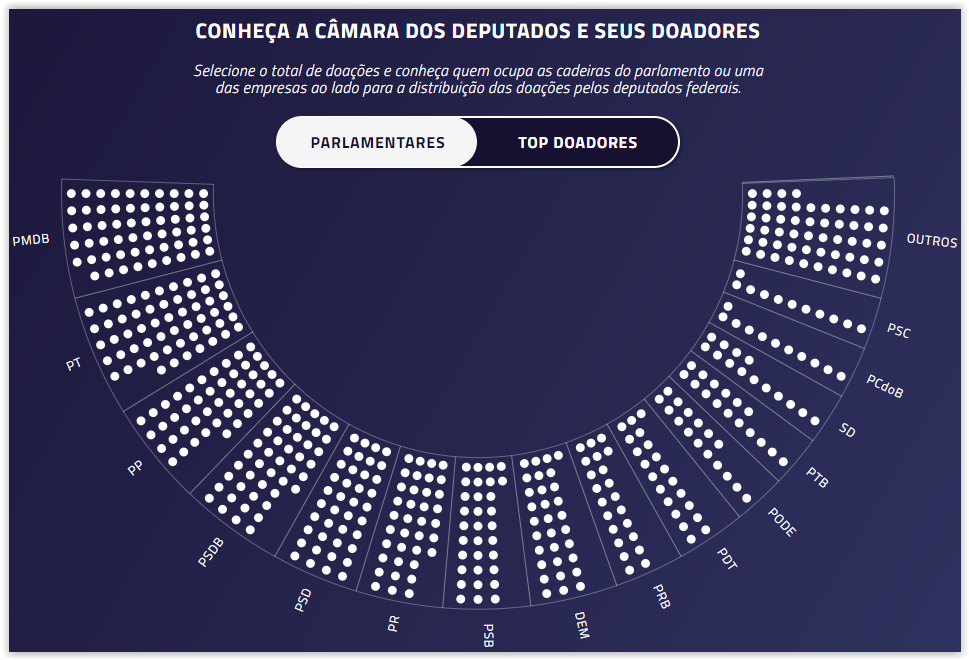


Figura 4 Mapa da Câmera

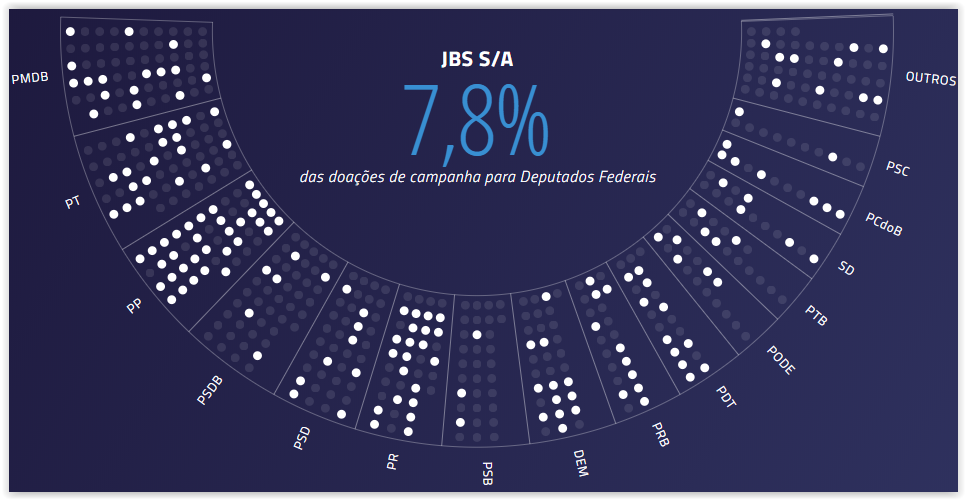


Figura 5 Doações de Empresas para campanha de Deputados Fendereis



Figura 6 Lista de Empresas Doadoras

O projeto vai além e mostra uma lista com as empresas que mais contribuem com as campanhas eleitorais de Deputados Federais, mostrando para quais debutados foi tem o apoio dessas empresas (Figura 5 e Figura 6). Os “top doadores” também são claramente apresentados junto com o que equivale o montante de sua contribuição (Figura 7).

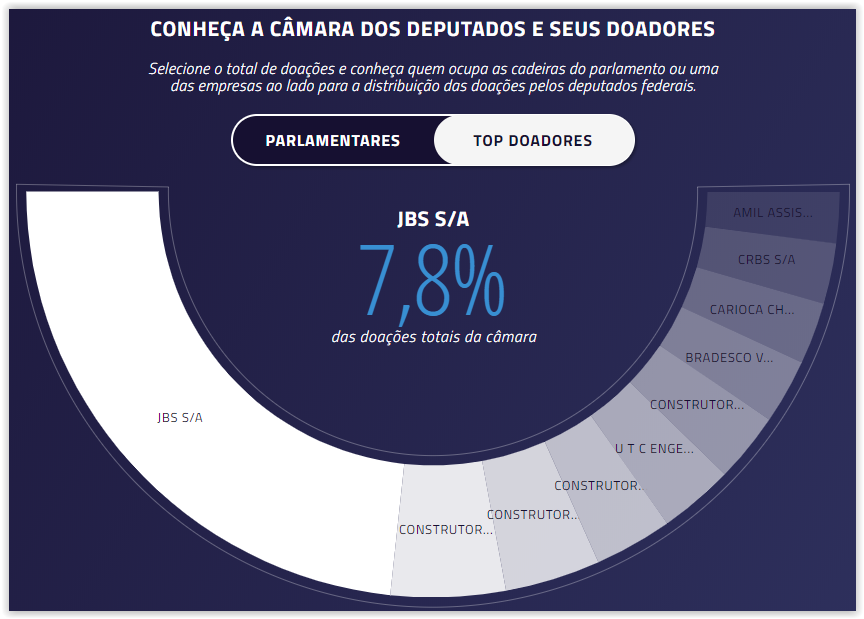


Figura 7 Top Doadores

É possível visualizar o montante que cada partido recebeu (Figura 8) e quanto cada estado venho a receber (Figura 9).



Figura 8 Quanto foi doado para cada partido?

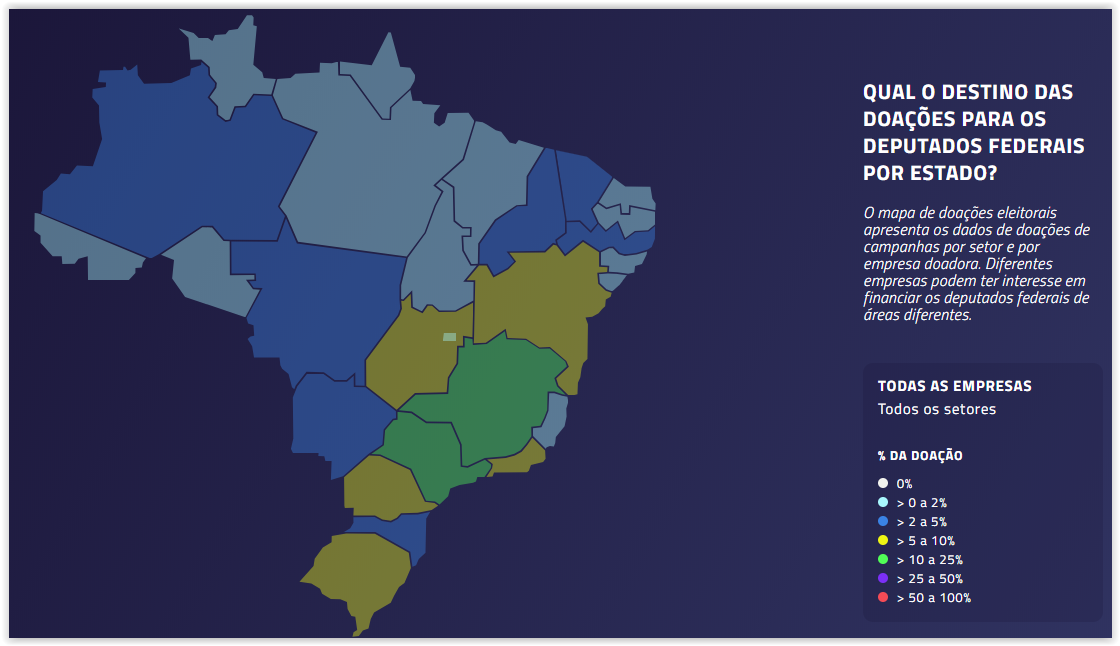


Figura 9 Qual o Destino das Doações para os Deputados Federais por Estado

**2.5 CONCLUSÃO**

Ao escolhermos os trabalhos relacionados as principais características que procuramos foram: envolver política como tema principal, utilizar dados públicos e de acesso da população como fonte para os dados de suas aplicações e representar esses dados em tela de forma gráfica.

Todos eles possuem a característica de representar os dados em tela no formato de gráficos estatísticos, incluindo nossa proposta. Alguns desses projetos apresentam gráficos com dados fixos, outros apresentam opções de filtros para modificar as informações do gráfico para uma consulta mais especifica.

Nosso projeto irá manter os dados atualizados através de extrações agendadas, diferente dos trabalhos “House of Cunha” [1] e “Quem me Representa” [2] que utilizam dados que foram coletados apenas uma vez, já que tinham como foco eventos políticos em específico.

Diferente dos outros trabalhos que encontramos, pretendemos disponibilizar a busca pelos dados pós analisados de nossas bases através de Serviços Web.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **House of Cunha** | **Quem me representa** | **Serenata de Amor** | **Câmara Transparente** | **FGApp** | |
| **Apresenta Informações através de gráficos** |  |  |  |  |  | |
| **Mostra informações através de representação estatística** |  |  |  |  |  | |
| **Interação e Filtros** |  |  |  |  |  | |
| **Uso de IA** |  |  |  |  |  | |
| **Disponibilidade de Serviços Web para consultas** |  |  |  |  |  |

Tabela 1 Comparativo de recursos em trabalhos relacionados

Conforme o andamento e desenvolvimento do trabalho de conclusão, outros trabalhos semelhantes serão pesquisados e utilizados de base para o desenvolvimento de nosso trabalho. Os novos trabalhos encontrados serão devidamente adicionados e referenciados.

# FUNDAMENTAÇÃO TEÓRICA

Esta seção apresenta os principais conceitos teóricos que dão base a esse trabalho. Logo, aborda o que é ciência de dados na Seção 3.1 e na sequência aborda brevemente na Seção 3.2 sobre experiência de usuário.

## CIÊNCIA DE DADOS

A ciência de dados é um campo de estudo em expansão, segundo análise realizada pelo LinkedIn há 6,5 vezes mais cientista de dados do que há 5 anos atrás segundo os dados apresentados [7]. A disciplina de ciência de dados tem caráter hibrido que se relacionando com análise estatística e ciências da computação [8], dessa forma se relacionando com disciplinas de banco de dados, mineração de dados e sistemas distribuídos [9], assim formando m conjunto de métodos analíticos destinados a extrair informações de dados.

Como descrito em [10], a sociedade mudou do analógico para o digital em apenas alguns anos. Isso gerou um grande impacto na forma que se fazem negócios e do jeito que nos comunicamos [11] Os dados são recursos vitais para muitas organizações, sendo assim a gestão e análise de dados cada vez se torna mais importante, e isso não diz respeito somente a organizações empresariais, mas também a áreas da educação, saúde e muitos outros setores [12]. Sobre a premissa que os dados são recursos vitais e podem expor novas informações sobre um conjunto dados após uma adequada análise e processamento, o presente trabalho, propõe se fundamentar e utilizar técnicas usadas pela de ciência de dados para analisar dados do setor público referente a cargos comissionados e funções gratificadas e suas relações com base partidárias.

### A multidisciplinaridade de Ciência de dados

Como já mencionado anteriormente Ciência de Dados é formada por diferentes disciplinas parcialmente sobrepostas. Os limites entre as diferentes disciplinas que formam a Ciência de Dados não são muito claro e parecem mudar conforme o tempo. A seguir uma breve explicação sobre as disciplinas que são englobadas pela ciência de dados.

* **Estatísticas.** Pode ser vista como a primeira forma de analisar dados. Considerando a estatística descritiva que pode ser usada para resumir dados de amostras usando noções como média, desvio padrão, frequência [9].
* **Mineração de dados.** Tendo sua origem com ligada a área de Banco de dados pode ser definida como análise de conjunto de dados geralmente muito grandes para encontrar novas relações e resumir os dados de uma nova maneira [9].
* **Aprendizado de Máquina.** Está disciplina se preocupa com construir programas que tenham capacidade de aprender e se adaptar a partir de um modelo e de suas entradas de dados [9].
* **Mineração de Processos.** Em conjunto com o aprendizado de máquina a mineração de processos busca confronto entre dados e dados de eventos (ou seja, comportamento observado) e modelos de processo (feitos à mão ou descobertos automaticamente) [9].
* **Analise preditiva.** É a prática de extrair informações de dados existentes define a fim de determinar padrões e prever resultados e tendências futuros. Para gerar previsões, as abordagens existentes de mineração e aprendizagem são aplicadas [9].
* **Banco de dados.** Os sistemas de gerenciamento de banco de dados apresentam dois principais propósitos: estruturar dados para que possam ser gerenciados facilmente e fornece escalabilidade e desempenho confiável. Dessa forma remove a responsabilidade de um programa administrar os dados. Banco de dados relacionais do tipo *SQL* (*Structured Query Language*) eram a norma, mas devido ao grande aumento do volume de dados os bancos de dados distribuídos do tipo *NoSQL* passaram ser fortemente utilizados [9].
* **Sistemas distribuídos.** Fornecem a infraestrutura para conduzir análises. Uma distribuição sistema é composto de componentes que interagem que coordenam suas ações para alcançar um objetivo comum. Computação em nuvem, rede e utilitários dependem de sistemas. Algumas tarefas de análise são muito grandes ou muito complexas para serem executadas único computador. Essas tarefas podem ser divididas em várias tarefas menores que podem ser executadas simultaneamente em diferentes nós de computação. A escalabilidade pode ser realizada compartilhando e / ou estendendo o conjunto de nós de computação [9].
* **Visualização e análise visual**. No final as pessoas precisam interpretar os resultados e orientar a análise. A análise depende muito do julgamento humano e da interação direta com os dados. As capacidades de percepção do sistema cognitivo humano podem ser exploradas usando as visualizações certas [13]. A análise visual, termo cunhado por Jim Thomas (1946–2010), combina técnicas de análise automatizada com visualizações interativas para um entendimento, raciocínio e tomada de decisão eficazes, com base em conjuntos de dados muito grandes e complexos [14].

Percorrendo brevemente diversas disciplinas que contribui para formar a Ciência de Dados, é sem dúvida um campo de estudo que envolve diversos vieses. Com isso em vista, é necessário agora entender como os dados essas áreas vão se sobrepor de forma a concretizar o conceito de ciências de dados.

### Dimensões da Análise de Dados

Quando se utiliza um grande volume de dados, é comum que o conjunto de dados não esteja apto a ser a processado ou analisado. Os dados brutos, na maioria das vezes não estão modelados de acordo com questões que o analista pretende responder. O Analista, terá um trabalho duro e deverá usar uma estratégia inteligente que deve ser executada com cuidado [12], o problema da qualidade de dados que comumente se relaciona com erros nos dados, dados incompletos ou dados faltando [8], podem ser enfrentando através de uso de mapeamento de dados que ajudará a desenvolver uma estratégia para modelar os dados de acordo com o problema que necessita ser solucionado

Para nortear a coleta de dados o cientista de dados pode se fazer as seguintes questões:

* (Relatório) O que aconteceu?
* (Diagnóstico) Porque aconteceu?
* (Predição) O que pra acontecer?
* (Recomendação) O que de melhor pode acontecer?

As questões, a pesar de simples fazem com o que cientista de dados se questione sobre a quais dados selecionar para responder o problema que deseja responder. Desse modo, se inicia a análise do conjunto de dados e se direcionar ao que se deseja inferir com os dados da análise.

* **Preparo de dados.** Para que seja possível realizar uma análise eficiente é necessário que dados usados tenham qualidade. É comum que os dados recém extraídos de suas respectivas fontes tenham erros, sejam incompletos ou tenham redundâncias de informações. Para evitar transtornos de possíveis erros uma pratica essencial é questionar a validade, a precisão a integridade e consistência dos dados. Selecionar a fonte da dados é a primeira etapa para construir uma base de dados para serem analisados. Segundo AKERKAR & SAJJA [8], sempre há um forte desejo de incluir todos os dados disponíveis, a máxima quanto mais melhor será válida. Mas isso pode ou não ser verdade. É necessário considerar que o dado atenda o problema a ser respondido.
* **Pré-processamento de dados.** considerando as fontes de dados selecionadas previamente, é necessário criar estratégias para extrair e combinar dados de diferentes fontes. O processo ELT (*extact, load and transformation*), que é um processo de *data* *warehousing*, estrutura essa etapa. A extração envolve a conexão com sistemas de origem selecionando e coletando o dado. Transformação é a execução de uma série de regras para transformar os dados extraídos em padrões formatados. A carga (*load*) significa importar os dados para uma infraestrutura de armazenamento. Identificar informações imprecisas, incompletas ou dados irracionais, assim como remover dados redundantes caracteriza a limpeza de dados. A limpeza de dados contribui para tornar o conjunto de dados mais coesos.
* **Visualização de dados.** É a disciplina dedicado aos princípios e métodos de tradução de dados para a forma visual [23]. Humanos são animais visuais. Nós absorvemos informações sensoriais de forma mais eficiente através da visão. Não é surpresa que a visualização de dados seja muito eficaz para extrair informações de dados [23]. Uma visualização é eficaz quando pode ser decodificada com rapidez e precisão pelo público - os pontos principais devem ser quase imediatamente identificados [23]. A regressão de dados pode pegar dois ou mais conjuntos de dados e determinar o nível de dependência e uma equação para correspondência. Este é o equivalente matemático de modelagem de dependência. Regressão pode determinar uma equação e calcular com que proximidade os pontos de dados correspondem à linha.
  + **Regressão de dados** pode pegar dois ou mais conjuntos de dados e determinar o nível de dependência e uma equação para correspondência. Este é o equivalente matemático de modelagem de dependência. Regressão pode determinar uma equação e calcular com que proximidade os pontos de dados correspondem à linha.
  + **Detecção de anomalias**: as técnicas de gráficos fazem desvios da norma prontamente aparente. Alguns desvios são erros que podem ser removidos do conjunto de dados. Outlier A detecção identifica quais pontos devem ser analisados ​​para determinar sua relevância.
  + **Modelagem de dependência**: muitas vezes, dois conjuntos de dados tenderão ou alternarão juntos porquê de alguma dependência. Outros relacionamentos menos óbvios podem ser descobertos pela dependência modelagem. As empresas podem monitorar fatores acessíveis (como o clima) prever menos fatores aparentes, como vendas de um determinado tipo de produto. Nos gráficos, um relacionamento positivo entre dois conjuntos de dados aparecerá aproximadamente como uma linha.
  + ***Clustering*:** à medida que os conjuntos de dados são apresentados, os analistas podem descobrir uma tendência para os dados aponta para agrupar em grupos. Isso pode revelar relações de dados de maneira semelhante à modelagem de dependência, mas para variáveis discretas.
  + **Classificação de dados:** uma maneira de usar dados paramétricos para classificar entidades, semelhante a agrupamento. Por exemplo, uma companhia de seguros pode usar dados de "vida cotidiana" para um cliente para determinar se o cliente está em risco ou não.

## EXPERIÊNCIA DO USUÁRIO

A experiência do usuário é a experiência que o produto cria para as pessoas que o usam no mundo real. Quando um produto está sendo desenvolvido, as pessoas prestam muita atenção ao que ele faz. A experiência do usuário é o outro lado da equação, muitas vezes negligenciado, como funciona. A Experiência do Usuário não é sobre o funcionamento interno de um produto ou serviço. Experiência do Usuário é sobre como ele funciona no exterior, onde uma pessoa entra em contato com ela [15].

Experiência do usuário, ou UX (*User* *eXperience*), é a experiência do consumidor ao usar seu produto ou serviço. A UX tem um papel tão importante hoje que mesmo profissionais de *marketing* precisam entender um pouco disso. Isso porque a UX trabalha com questões funcionais e práticas, mas também com todo o lado emocional da experiência [16].

Criada em 2011, a ISO 9241-210 define a experiência do usuário como "as percepções e reações de uma pessoa que resultam do uso ou utilização prevista de um produto, sistema ou serviço." De acordo com a definição da ISO, experiência do usuário inclui todas as emoções, crenças, preferências, percepções, respostas físicas e psicológicas, comportamentos e realizações do usuário que ocorrem antes, durante e após o uso.

## CONSIDERAÇÕES da seção

Os aspectos relatados na Seção 3 serão utilizados para guiar a execução deste trabalho. Os campos de estudo como ciência de dados, mineração de dados, armazém de dados são intrinsecamente relacionados. Esses campos dão norte para o que se trata de análise captura de dados. Já as outras áreas citadas nessa seção guiarão o processo de criação de interface, de forma a oferecer ao usuário uma ferramenta fácil de ser utiliza. De mesmo modo que guia a maneira de exibir os dados, sempre buscando deixar as informações clara e de fácil.

Os presentes tópicos brevemente explicados nessa seção serão explorados e aprofundados. Outras áreas ainda não citadas poderão ser incluídas ao longo deste trabalho de forma a provisionar conhecimento.

# PROPOSTA

Esta seção descreve nossa proposta de trabalho de conclusão de curso. Na Seção 4.1 justificamos a necessidade da criação de nossa aplicação. A Seção 4.2 apresenta o público que utilizara a ferramenta. A Seção 4.3 informa onde será disponibilizado o acesso a nossa aplicação. A Seção 4.4 descreve como a ferramenta deve continuar a longo prazo. A Seção 4.5 apresenta os principais objetivos a serem entregues pelo trabalho. A Seção 4.6 descreve a ferramenta em si de forma breve. A Seção 4.7 evidencia os possíveis riscos que podem comprometer o desenvolvimento do trabalho.

## Necessidade

No Brasil, o cenário político passa por uma série de escândalos de corrupção. Isso ressalta a necessidade de maior transparência no que diz respeito aos dados públicos afim de se fiscalizar e avaliar as ações tomadas pelos nossos representantes no governo. Em virtude disso, esse trabalho busca analisar os dados disponibilizados pelo Portal da Transparência e TSE em busca de novas informações e relações perante servidores que ocupam cargos comissionados.

O número de servidores que ocupam cargos ou funções de comissão é considerável, sendo assim, compreender a movimentação desses se torna importante. Isso se justifica, pois é considerado normal na política brasileira que muitos políticos usufruem cargos e votos como moeda de troca. Muitos trocam de cargo para que possam influenciar em votações de acordo com seus interesses. Ou até mesmo ganham cargos melhores com objetivo de garantir votos no Plenário da Câmara contra ou a favor de seus grupos e coligações.

## Público-alvo

Os dados que os cidadãos têm acesso hoje são apresentados de forma simples sem um processamento ou esses dados não apresentam uma relação. Se tratando de uma ferramenta gratuita e online, o público alvo se torna o usuário de internet que tem interesse sobre informações sobre dados públicos que foram explorados nesse trabalho.

## Disponibilidade

Por se tratar de uma ferramenta online e projetada para atender múltiplas plataformas (móvel ou computador) o serviço estará sempre disponível ao usuário final. As bases de dados também serão constantemente atualizadas a fim de sempre mostrar informações reais para o usuário final.

## Legado

Com base em conhecimentos adquiridos durante a graduação e apresentados na Seção 3 e em trabalhos relacionados citados na Seção **Error! Reference source not found.**

Buscamos dar importância a ciência de dados e a sua aplicação direcionada ao campo político social de modo a promover maior entendimento sobre informações. Dessa forma disponibilizar uma ferramenta que facilitasse o entendimento desses de dados.

## Objetivos

Criar uma ferramenta que identifique movimentação de servidores que ocupam cargos comissionados, funções gratificadas e funções similares. Dessa formam, mostrar a evolução do número de servidores quem ocupam uma dessas funções com passar do tempo. A distribuição de servidores com FG (Função Gratificada) pelo país. A relação de servidores com partidos políticos. movimentação de cargos e a movimentação entre os cargos assim como descobrir a qual partido esse servidor é vinculado são os principais objetivos que esse trabalho propõe.

Baseando-se em métodos computacionais do campo da ciência de dados e de iteração de usuários se tem como objetivo exibir as relações descobertas de forma simples e clara para o usuário final, assim como possibilitar o usuário uma simples interação com o sistema.

## Ferramenta

A ferramenta deverá funcionar de forma semiautônoma sempre buscando atualização dos dados e executando processamentos de forma a sempre dispor de dados atualizados para o usuário. Um usuário administrador poderá interagir com o sistema para eventual correção do de algum erro ou ruptura no processo de obtenção e análise de dados.

A ferramenta, consiste em capturar dados públicos referentes a funções atribuídas a servidores públicos da esfera federal com o objetivo de processá-los a fim de expor novas informações descobertas por meio da análise de dados. As informações descobertas serão apresentadas em um aplicativo *web* que será criado utilizando os conceitos de UX para que o usuário final possa ter a melhor usabilidade. Sendo assim, poderá ser acessada através dos navegadores mais utilizados atualmente, podendo ser acessada tanto de computadores pessoais quanto de dispositivos móveis. Por fim, a ferramenta possibilitará o uso de diversos filtros, poderá também ser usada por jornalistas como uma fonte de dados quando estes desejarem utilizar essas informações para reforçar o embasamento de notícias e textos.

## O risco do projeto

O projeto se inicia com dois riscos iniciais.

* **Fonte da dados.** Sendo o Portal da Transparências e o site do TSE fontes de dados que este trabalho utiliza, há o risco dessas fontes pararem de atualizar usas bases de dados assim impossibilitando a captura dos mesmos para serem utilizados na análise.
* **Relacionamento entre os dados.** Se busca encontrar relações entre os servidores, (dados do Portal da transparência) com dados de filiados a um partido (TSE). Com isso se espera se mapear servidores com partidos. Conseguindo assim mensurar quantos servidores são relacionados a um partido.

# Descrição do sistema

A presente Seção tem como objetivo descrever o sistema. Apresentado, inicialmente as funcionalidades (Seção 5.1) de interesse do usuário final, apresentado na Seção 6.2. Para se atingir as funcionalidades descritas pela Seção 5.1, são apresentas as funcionalidades de *Back-End* (Seção 5.2), que constituem o esforço necessário para que o sistema atinja s resultados final.

## Funcionalidades

Buscando trazer maior transparência e clareza de informações sobre Cargos Comissionados sistema busca apresentar os dados que foram capturados e processados de forma mais simples.

**Distribuição de Cargos Comissionados.** Apresentar ao usuário uma forma clara de ver a distribuição de cargos comissionados por estados brasileiros. Assim, como o númerode ocupantes em cada tipo de função gratificada que existe.

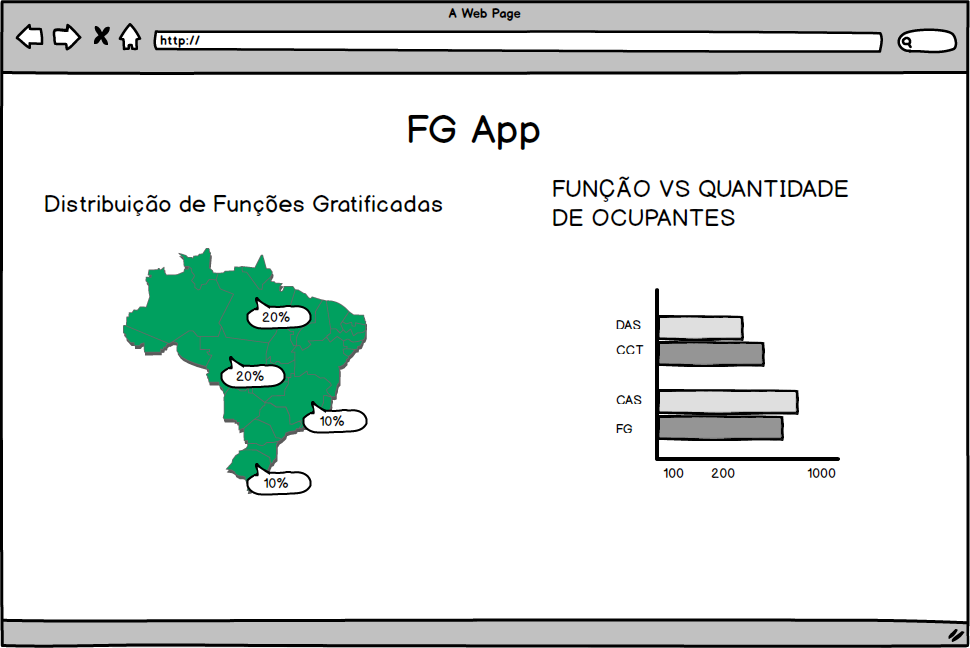


Figura 10 Distribuição de Funções Gratificadas no País

* + **Top servidores que trocaram de função.** Através da análise dos dados que foram extraídos (seção ...), se identifica o período de tempo de cada servidor esteve vinculado a uma função. Assim, como a função em que o servidor se encontra no no momento atual. Logo ao usuário final se apresenta uma lista de servidores que trocaram de cargo recentemente, informando qual a função atual e qual foi a função anterior que o servidor ocupava. Figura 11.

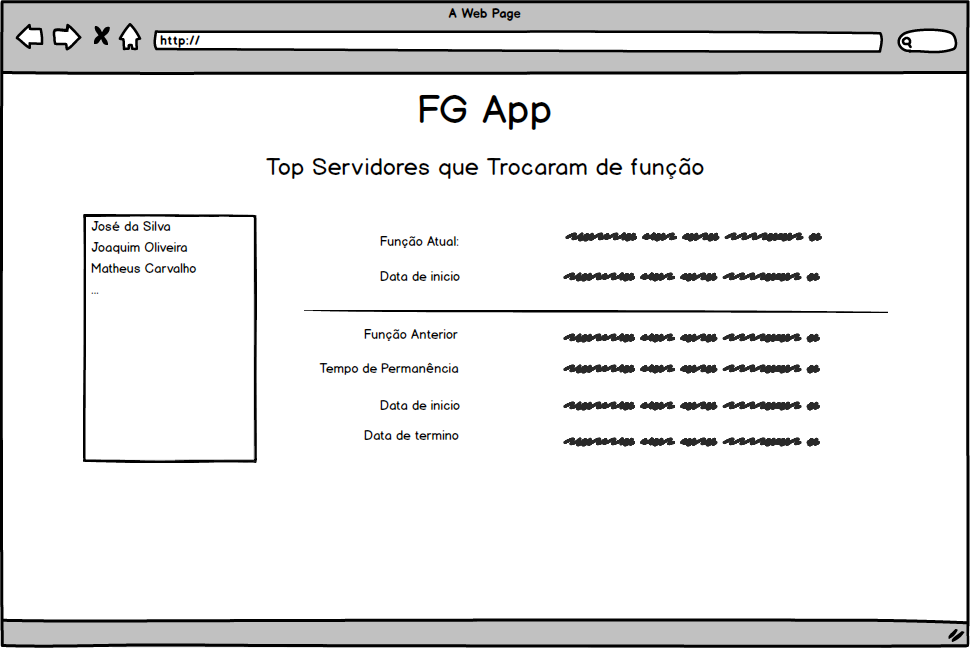


Figura 11 Top Servidores que trocaram de Função

* + **Partido Político e Servidores.** Uma das principais analises que se tem interesse em mostrar ao consumidor final, é a descobrir o percentual de servidores, que ocupam um cargo comissionado, os quais são filiados a um partido político. Através dessa relação é possível expandir a compreensão da influência que as bases partidárias exercem sobre cargos desse tipo. A Figura 12, busca demonstrar como se deseja mostrar essas informações. A visualização através de gráficos do formato pizza é uma forma simples de mostrar o valor percentual.

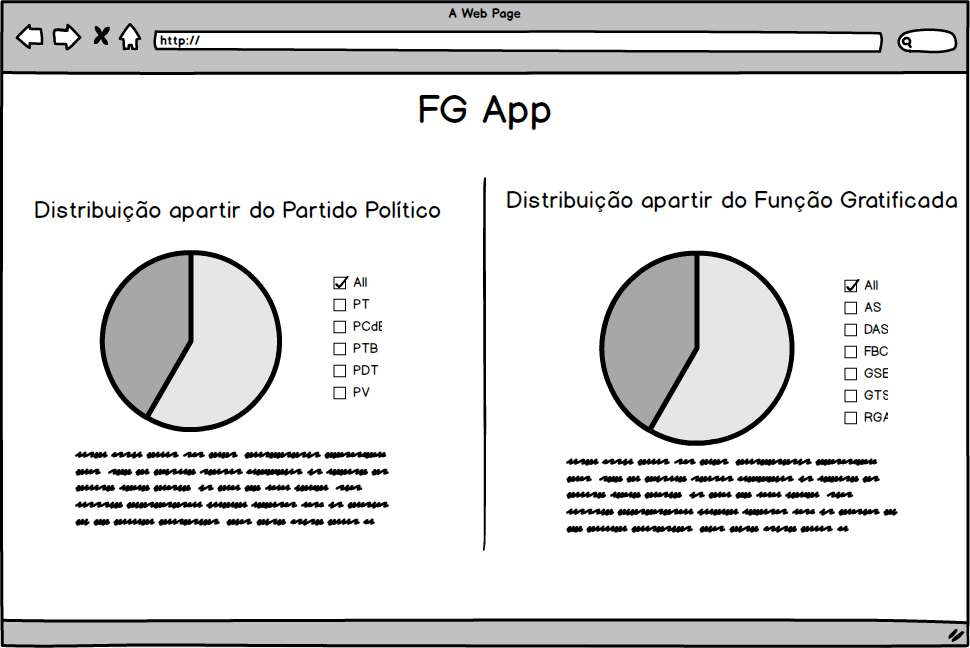


Figura 12 Partidos políticos vs servidores

* + **Evolução dos Cargo.** Essa funcionalidade se compromete em mostrar a evolução, ao longo do tempo, perante a quantidade em número de servidores. Sendo possível filtra por cargos ocupados e por estado. A Figura 13, demostra um projeto inicial da tela que representará essa função.

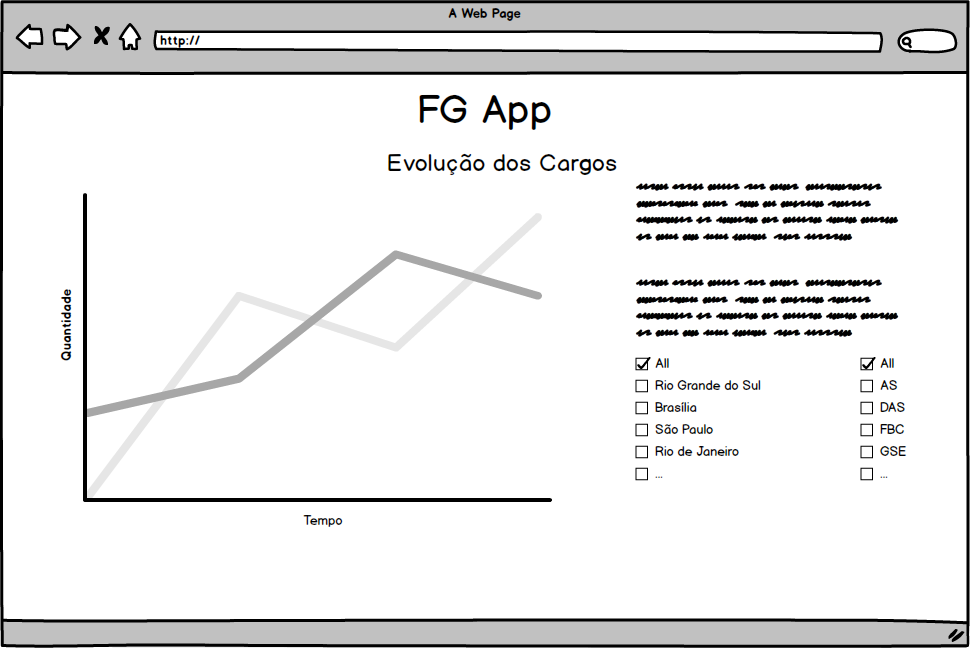


Figura 13 Evolução do nº de servidores

## back-end do sistema

Essa Seção, contém a explicação das rotinas que compõe o sistema para atingir os objetivos descritos na seção anterior. Portanto, ela não apresenta iteração por parte do usuário final, mas somente por parte do usuário do tipo administrador do sistema.

A primeira parte dessa sessão apresenta as fontes de dados selecionadas para o desenvolvimento para construção desse trabalho. Sendo o site do Portal da transparência e o do Tribunal Superior Eleitoral, os retentores das informações sensíveis ao contexto dessa pesquisa, será explicado brevemente onde e como os dados alvos se apresentam.

Após isso, uma breve explicação sobre como os dados serão manipulados e salvos, afim de atingir-se o objetivo final já descrito. Informações mais específicas sobre a arquitetura do sistema serão abordados na Seção 6.

A origem dos dados que são utilizados por esse trabalho é de origem publica governamental. As fontes são o Tribunal Superior Eleitoral no que diz respeito a partidos políticos e seus filiados e o Portal da Transparência que disponibilizada informações sobre servidores e cargos que esses ocupam. A seguir uma breve explicação sobre as fontes de dados.

* + **Tribunal Superior Eleitoral.** O TSE órgão máximo da Justiça Eleitoral, exerce papel fundamental na construção e no exercício da democracia brasileira. Suas principais competências estão fixadas pela Constituição Federal e pelo Código Eleitoral (Lei nº 4.737, de 15.7.1965) [17]**.**  O site do TSE contribuiu para esse trabalho apresentando informações sobre partidos políticos e seus filiados, tais informações estão disponíveis para download. Para tal é necessário selecionar o partido, o estado que se deseja e apor clicar no botão salvar. Portanto é notável que no site do TSE não é possível observar esses dados diretamente na página (Figura 14). No site do TSE também não apresenta informações sobre quando foi a última atualização da relação, contudo, pelas informações contidas nos arquivos baixados os dados parecem ter uma atualização mensal.

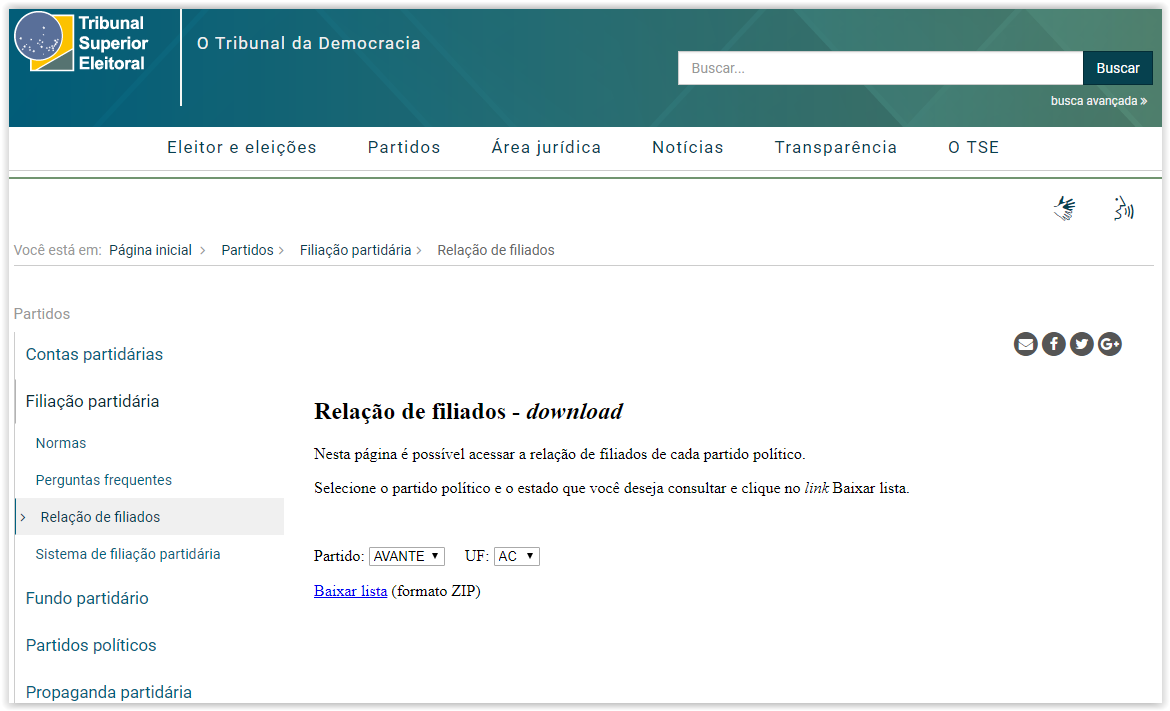


Figura 14 Site do TSE

* **Portal da Transparência.** As informações existentes no Portal referem-se ao Poder Executivo e à esfera federal. A ferramenta publica, ainda, dados sobre assuntos transversais ou que estejam relacionados à função da maioria desses órgãos [18]. Dessa forma o Portal traz informações sobre servidores e funções gratificadas. Além disso, apresenta formas de se visualizar os dados diretamente no próprio portal. A Figura 15 mostra a primeira etapa do processo de consulta de servidores.

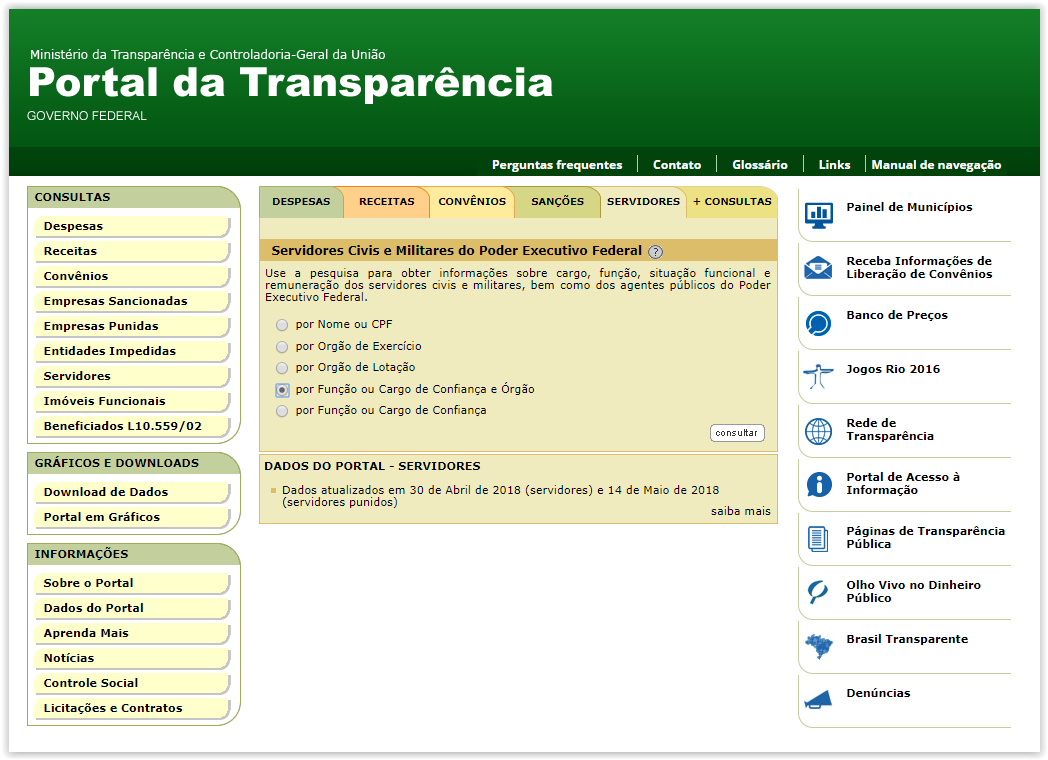


Figura 15 Consulta de Servidores

Focando nos servidores que ocupam função ou cargo de confiança o próprio portal já mostra a separação por cargos e o número de servidores que tem ocupado essas posições (Figura 16). Navegando pelos através da seleção de uma função ainda é existe o grupamento por nível da função antes de se alcançar a lista de servidores (Figura 17).



Figura 16 Lista de funções e número de ocupantes



Figura 17 Lista de servidores

Ao selecionar um servidor se encontra uma sucessão de informações como é possível observar na Figura 18, entretanto para visualizar a remuneração desse servidor é necessário se direcionar outra página (Figura 19), onde cada mês é relatado numa aba diferente.

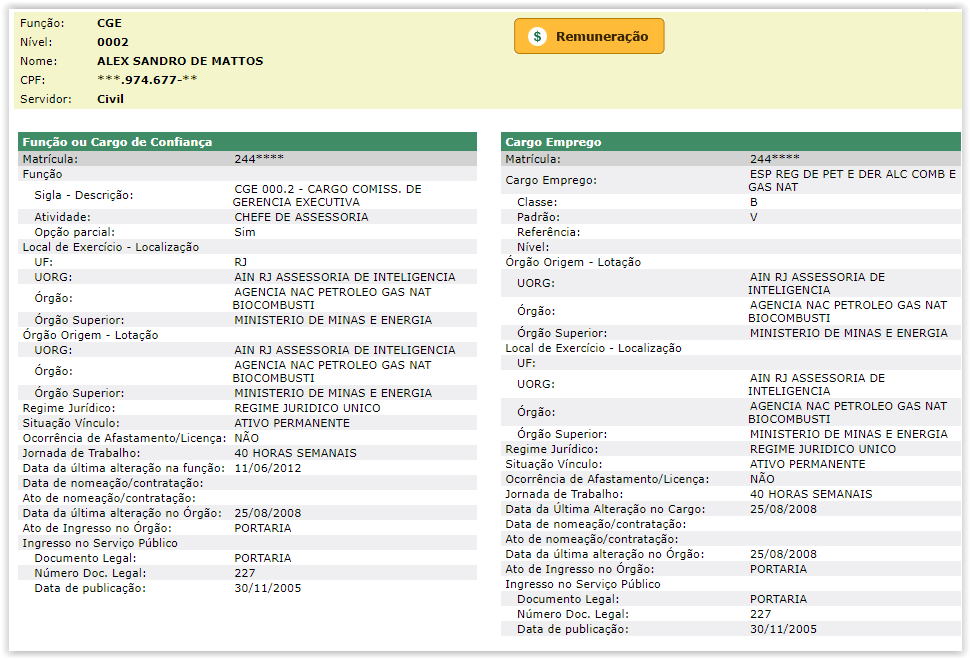


Figura 18 Detalhes sobre um servidor

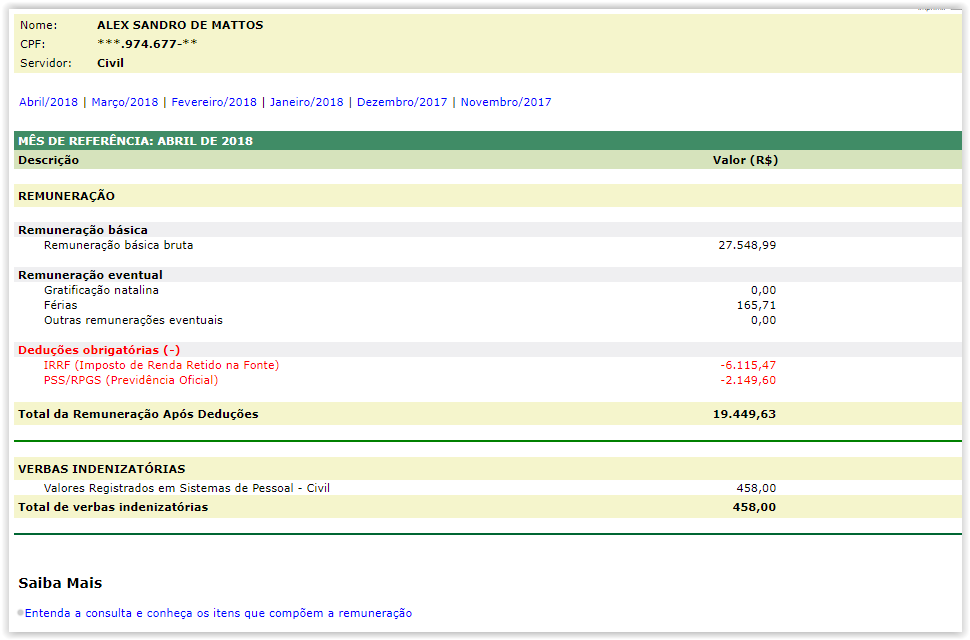


Figura 19 Remuneração de um servidor

Portal da transparência mesmo apresentando os dados de servidores diretamente no site também disponibiliza a opção de download dessas informações. Os arquivos são organizados por ano e mês ainda podendo ser fragmentado em registros de servidores civis. De ambas as fontes é possível fazer o download dos dados, dessa forma os dados são extraídos em formato CSV, a Figura 20 ilustra como os dados num formato CSV é disposto.

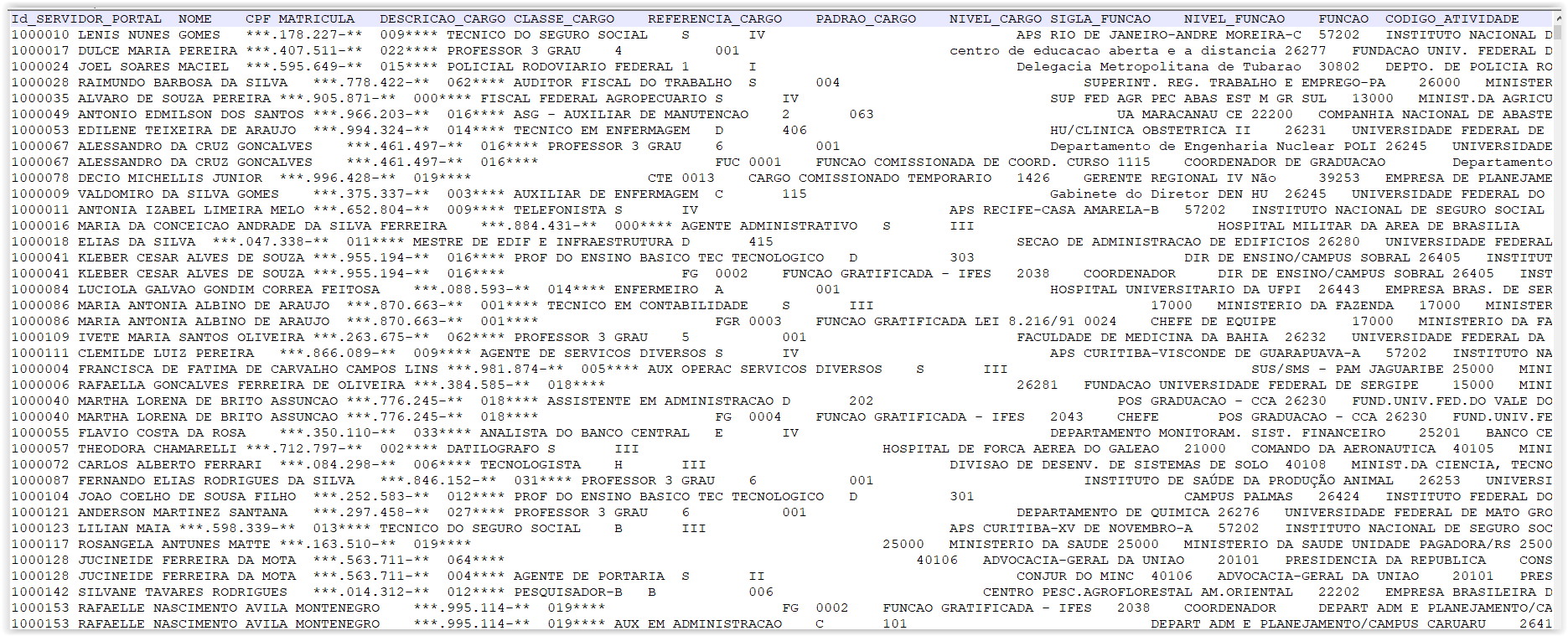


Figura 20 Dados em formato CSV

O volume de dados extraídos é significativo, além disso, para ser viável fazer uma análise dos dados é necessário que estes estejam coesos entre si. É relevante ressaltar que informações que não contribuem com a questão que se busca resolver não serão mantidos.

Os dados do Portal da Transparência serão analisados a partir de janeiro de 2013 até os dados mais recentemente disponibilizados que datam de abril de 2018. O modelo de dados se modificou nesse intervalo de tempo, sendo assim se exige uma maior atenção em como processar essas informações, novos atributos foram disponibilizados e alguns campos sofreram alteração de nome. Os dados do TSE, com a análise já realizada não apresenta variação no quesito atributos que compõe o conjunto de informações extraídas, portanto se torna de manipulação mais simples

Como já mencionado, será utilizado informações de anos anteriores, logo, o processo de obtenção de dados será realizado em duas etapas. Uma para obter dados de anos anteriores e outra para futura atualizações de acordo a disponibilidade.

A explicação de mais detalhado sobre o processo de extração de dados e demais atividades serão explicadas na seção seguinte, que apresenta caráter mais técnico definindo a arquitetura da aplicação.

# Arquitetura do sistema

A Seção apresenta definições e conceitos de arquitetura que serão utilizadas para o desenvolvimento do software proposto.

## Modelo de arquitetura em camadas

Uma arquitetura em camadas é projetada como uma hierarquia de processos cliente servidor que minimiza as interações entre camadas. Cada camada age como um cliente do módulo acima dela e também como um servidor daquele abaixo dela em uma arquitetura em camadas [19]. Sobre esse contexto, esse trabalho tem como objetivo isolar em módulos o funcionamento do sistema. A arquitetura que será exposta a seguir será implantada num ambiente de computação em nuvem *o Amazon Web Services (*AWS) provido pela Amazon.

* **Computação em nuvem.**  Podendo ser entendido como computação em nuvem, é a entrega sob demanda de poder computacional, desde armazenamento de banco de dados, aplicações ou outros recursos de TI por meio de uma plataforma de serviços de nuvem via internet com uma definição de preço conforme o uso [20].

Usando um ambiente de nuvem para implantar o projeto, definimos as camadas que serão implantadas como: camada de extração, camada de pré-processamento, camada de analise, camada de serviços, camada de usuário, que serão explicadas na sequência.

* **Camada de Extração.** A camada de extração tem como objetivo obter os dados originais disponibilizados pelas fontes selecionadas. Entretanto, após a obtenção dos dados que são arquivos do formato CSV, os dados são importados para um banco de dados a fim de promover maiores recursos sobre a manipulação dos dados. Um controlador de execução é definido como responsável por controlar as atividades de extração de dados e pela importação. Qual quer problema que ocorrer nessas etapas o administrador do sistema será notificado. Figura 21.

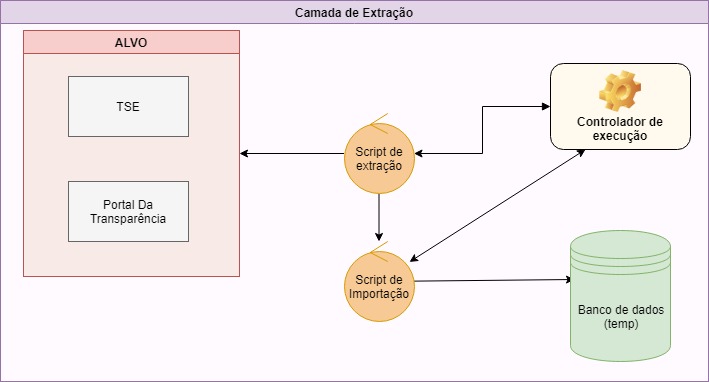


Figura 21 Camada de extração

* **Camada de pré-processamento.** Nessa etapa, mostrada pela Figura 22, os dados obtidos pela camada de extração serão limpos e organizados de forma a possibilitar uma análise futura. A camada anterior não se preocupava se os dados apresentam alguma anomalia ou não. Portanto, a camada de pré-processamento será responsável por preparar todas as informações para serem posteriormente analisadas, logo, atributos que sofreram alterações ao longo dos anos serão ressignificados a fim de garantir uniformidade do conjunto de dados, assim como as informações que não contribuem com o resultado pretendido serão removidos para não ocupar espaço desnecessário no banco de dados.

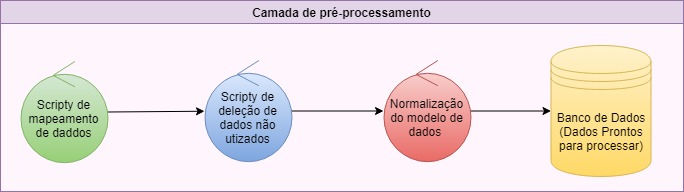


Figura 22 Camada de pré-processamento

* **Camada de análise de dados.** Tendo os dados já organizados em um modelo uniforme, essa camada representada pela Figura 23 tem como objetivo explorar a relação entre os dados disponíveis afim de revelar novas informações. A análise se caracteriza por rotinas cíclicas que sempre buscam descobrir novas informações, mas também avaliar a relevância dos resultados e sua coerência.

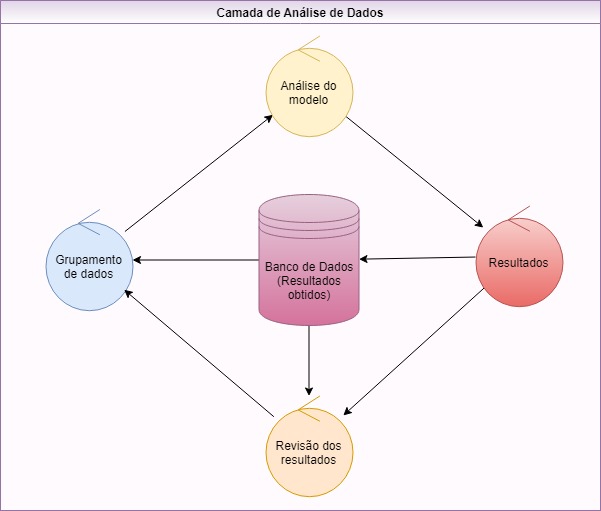


Figura 23 Camada de análise de dados

* **Camada de serviços.** A camada de serviços disponibiliza APIs que a interface do usuário usa para obter dados e resultados da análise.
* **Camada de Usuário.** A camada final do sistema onde o usuário interage. Nessa camada o usuário por intermédio da interface gráfica pode selecionar os dados e informações que deseja visualizar. O que o usuário selecionar acionará uma chamada a camada de serviço que retornara os dados necessários para gerar a visualização de dados e informações.

O que foi descrito nas subseções acima, tem como objetivo isolar funcionalidades de modo que outros módulos possam ser futuramente agregados ao presente sistema.

## Atores dos Sistema

Os papéis e ações já citadas ao longo desse trabalho, como usuário final e administrador, são ilustrados, nessa seção, através de diagramas UML de caso de uso. O diagrama de acaso de uso procura possibilitar a compreensão do comportamento externo do sistema, tentando apresentar o sistema por intermédio de uma perspectiva de usuário [21]. Os diagramas ilustrados pelas Figura 24 e Figura 25 se refere as atividades que cada participante do sistema pode realizar.

O usuário final, é para quem se destina o sistema proposto. Sendo assim o usuário final tem capacidade de interagir pelo sistema consultando as funcionalidades descritas na Seção 5.1.

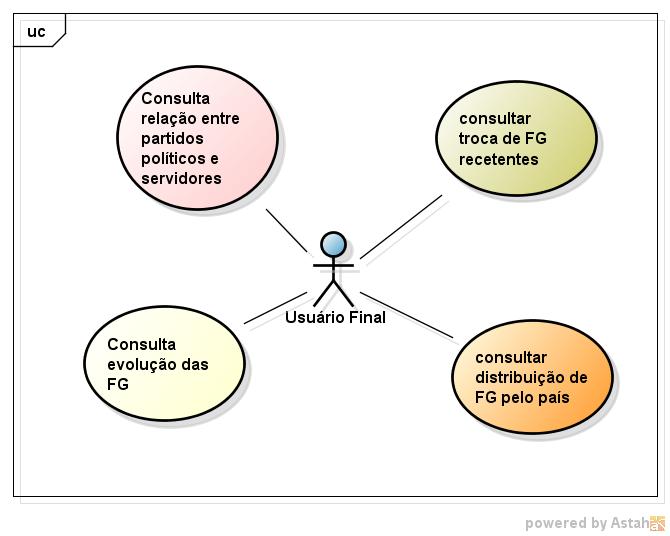


Figura 24 Diagrama de caso de uso – Usuário

O usuário administrador tem o controle total sobre o sistema, podendo interferir e interagir com qualquer camada e processo citados pela Seção **Error! Reference source not found.**. O administrador, deverá autenticar via console de administração do AWS para executar suas funções.

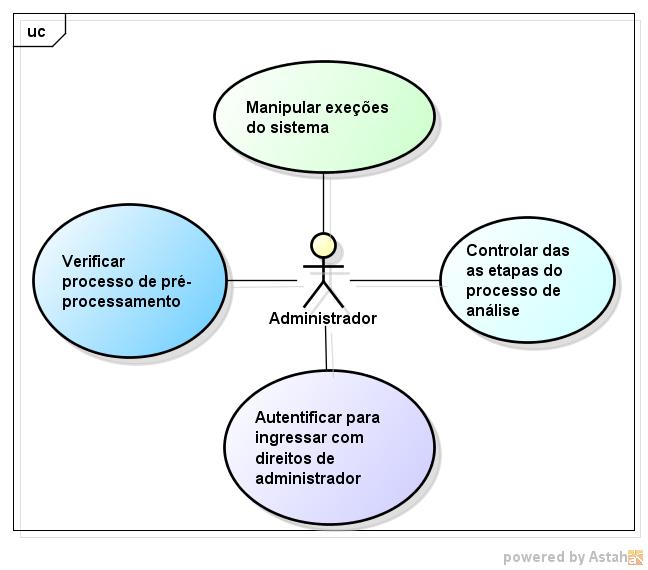


Figura 25 Diagrama de caso de uso – Administrador

## Diagrama de atividades

Utilizando recursos da UML, se apresenta nessa Seção os diagramas de atividade que ajudam a modelar o sistema. A modelagem de atividades enfatiza a sequência e condições para coordenar comportamentos de baixo nível. Dessa forma o diagrama de atividades é o diagrama com maior ênfase ao nível ênfase ao nível de algoritmo [21].

Este diagrama, pode modelar atividades de um método, algoritmo; ou ainda, um processo de nível de sistema. Uma atividade é composta por um conjunto de ações, ou seja, passos necessários para que a atividade seja concluída [21].

Os diagramas que estão na sequencia desse trabalho estão estruturados em nível de atividades exercidas pelas camadas descritas na Seção 6.1. Alguns diagramas representam uma única camada já outros apresentam atividades que cobrem mais de uma camada conforme a complexidade das atividades.

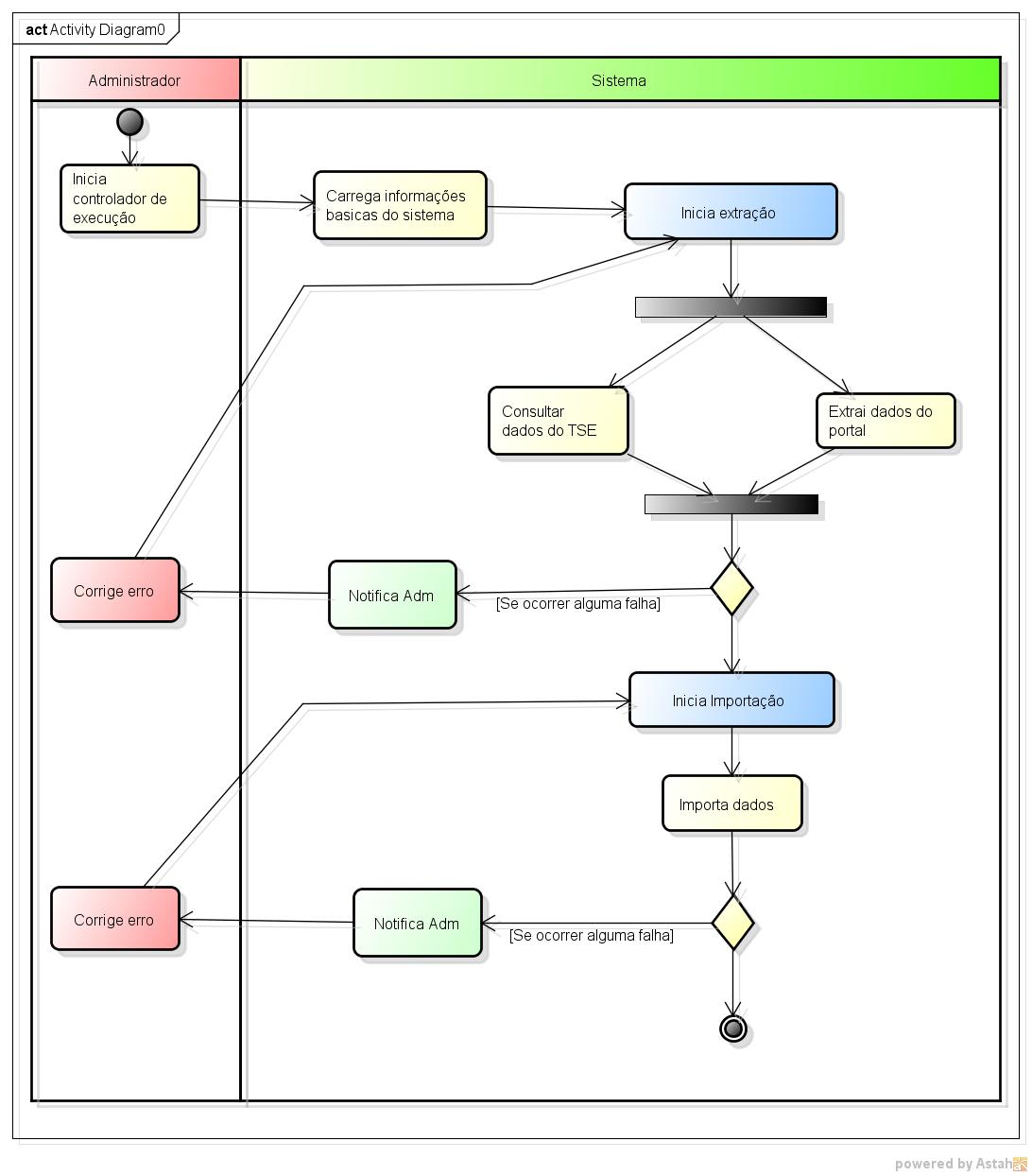


Figura 26 Diagrama de Atividade Camada de Extração

A Figura 26, apresenta as atividades que compõe a primeira camada do sistema – anteriormente já descrita. Sobretudo, o diagrama deixa explicito a atividades de responsabilidade do usuário administrador e do sistema. Nessa etapa não há nenhuma participação do usuário final.

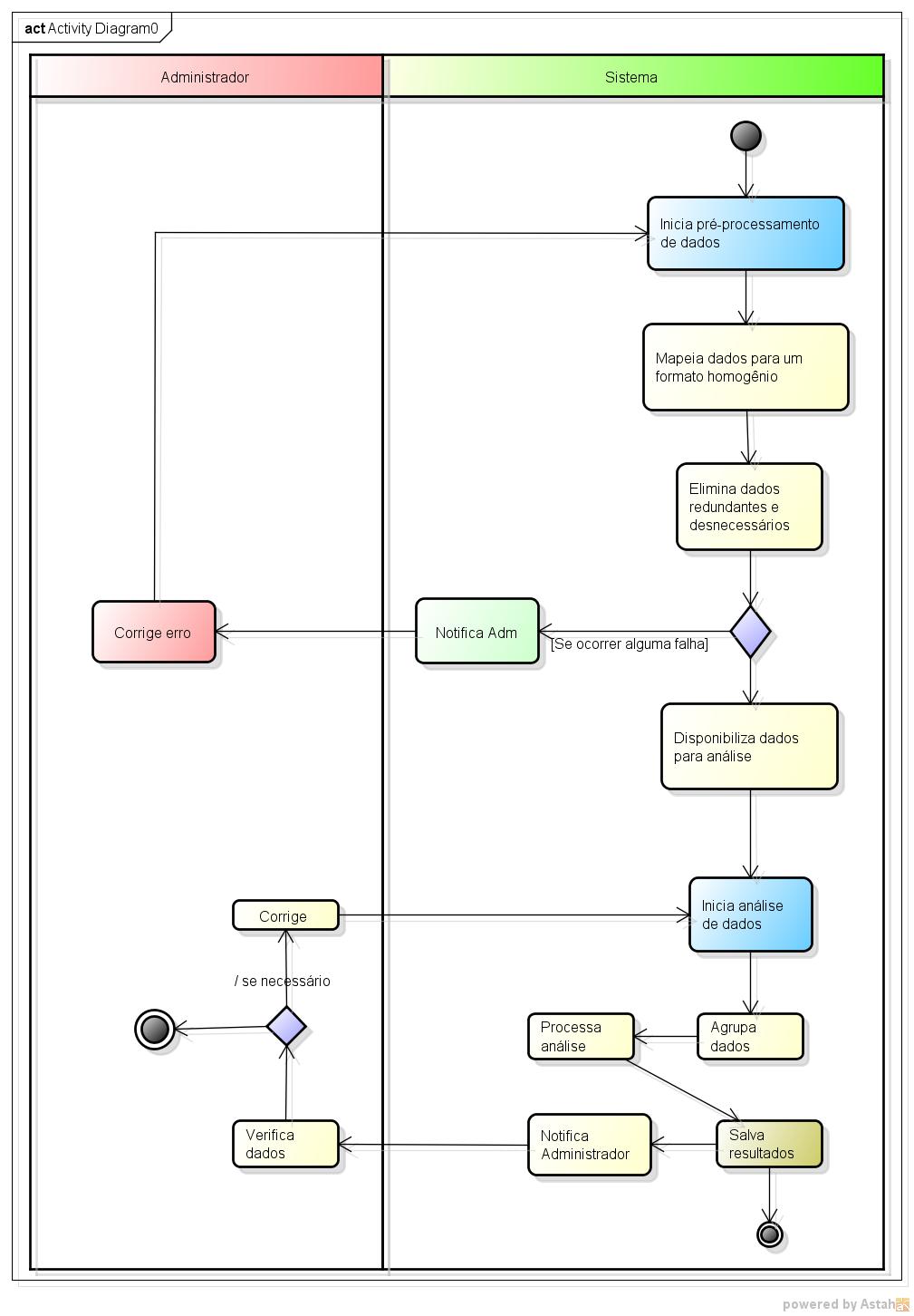


Figura 27 Diagrama de Pré-Processamento e de Análise de Dados

O diagrama representado pela Figura 27, apresenta atividade relativas a camada de pré-processamento e de análise de dados. Nessas camadas, o usuário administrador age quando ocorre falhas no sistema e confere os resultados após a análise de dados ser concluída.

Por fim, a diagrama mostrado pala Figura 28, mostra as ações em alto nível realizadas pelo usuário final desde o acesso ao sistema até o momento que ele deixa o sistema. As ações apresentadas pelo sistema é uma abstração das respostas que a cama da de serviço.

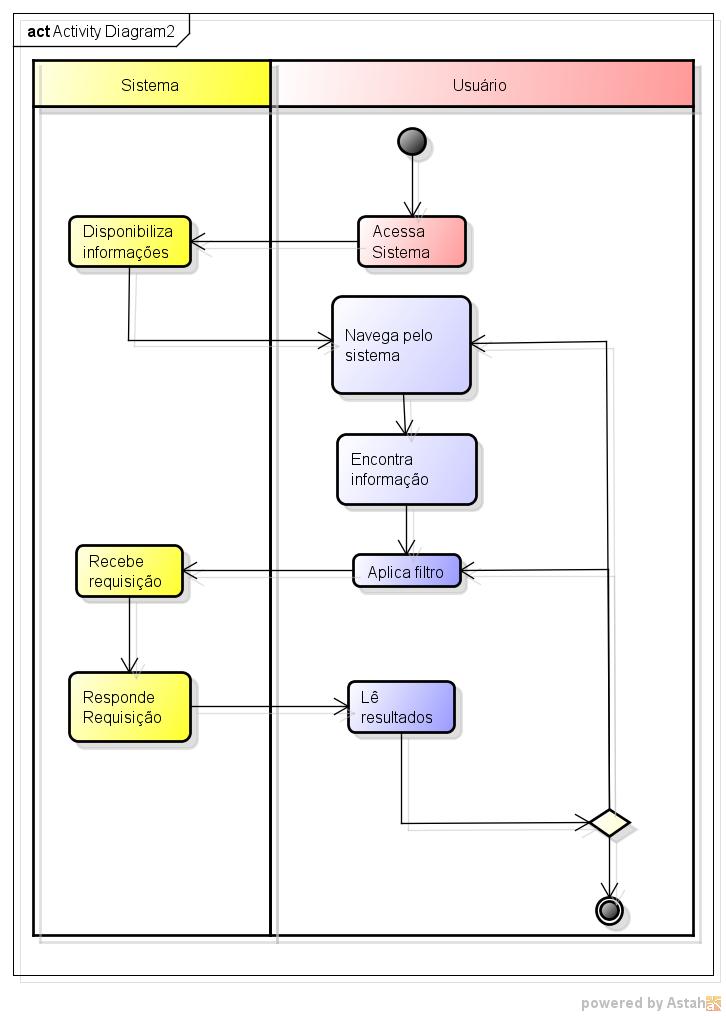


Figura 28 Diagrama de Atividade - Serviços e Usuário

Na Seção 6, se apresentou a arquitetura do sistema - é uma das etapas fundamentais para estruturar o sistema. Desse modo, foi aprofundado um pouco mais as características e funcionalidades do sistema. Os diagramas apresentados facilitam a compreensão das principais estruturas que compõe o sistema e as principais rotinas e atividades.

# CRONOGRAMA

Aqui é mostrado o cronograma planejado para a segunda etapa do Trabalho de Conclusão (TC2). O cronograma está organizado com as tarefas a serem executadas e como marco final, a finalização do documento escrito e do software desenvolvido e funcional. Além dos itens listados a baixo, estão previstas reuniões semanais com o orientador, visando buscar um alinhamento do trabalho realizado.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Atividades /Intervalo de dias | Agosto | | Setembro | | Outubro | | Novembro | | Dezembro |
| 01-15 | 16-30 | 01-15 | 16-30 | 01-15 | 16-30 | 01-15 | 16-30 | 01-05 |
| Configuração do ambiente |  |  |  |  |  |  |  |  |  |
| Desenvolvimento da camada de extração |  |  |  |  |  |  |  |  |  |
| Desenvolvimento do modulo de pré-processamento de dados |  |  |  |  |  |  |  |  |  |
| Desenvolvimento do modulo de análise de dados |  |  |  |  |  |  |  |  |  |
| Desenvolvimento da camada de serviços |  |  |  |  |  |  |  |  |  |
| Desenvolvimento do modulo de visualização de dados |  |  |  |  |  |  |  |  |  |
| Testes do sistema |  |  |  |  |  |  |  |  |  |
| Ajustes e correções de defeitos |  |  |  |  |  |  |  |  |  |
| Atualização do material teórico |  |  |  |  |  |  |  |  |  |
| Entrega do Volume 2 |  |  |  |  |  |  |  |  |  |
| Apresentação |  |  |  |  |  |  |  |  |  |

Tabela 2 Cronograma de Atividades

**Configuração do ambiente:** Criação de um ambiente para a implementação da aplicação. criação de um repositório de código e sistema de versionamento. Criação de ambiente para publicar versões de homologação do software.

**Desenvolvimento da camada de extração:** Implementação do ferramentas para extração automática de dados públicos disponibilizados em sites do governo.

**Desenvolvimento do módulo de pré-processamento de dados:** Análise de quais dados serão utilizados e como devem ser armazenados, e desenvolvimento de módulo que transformem os dados extraídos previamente para os desejados.

**Desenvolvimento do módulo de análise de dados:** Observar e explorar as possíveis relações entre os dados disponíveis. Implementação de módulo para encontrar agrupamentos e padrões nos dados coletados e processados.

**Desenvolvimento da camada de serviços:** Criação da camada de serviços que será disponibilizada para uso externo e também será utilizada pela nossa aplicação web.

**Desenvolvimento do módulo de visualização de dados:** Desenvolvimento da aplicação web com suas ferramentas visuais para interação com o usuário final.

**Testes do sistema:** Testes Finais da Solução Testes a serem executados sobre o produto final acabado.

**Ajustes e correções de defeitos:** Correção de problemas encontrados nos testes realizados e ajustes para uma versão final.

**Atualização do material teórico:** Novas alterações e resultados alcançados pelo grupo no desenvolvimento do trabalho final.

**Entrega do Volume 2:** Entrega do Volume 2 do Trabalho de Conclusão para avaliação.

**Apresentação:** Apresentação do mesmo frente à banca avaliadora.

# REFERÊNCIAS

x

|  |  |
| --- | --- |
| 1. | LABORÁTORIO ANALYTICS - UNIVERSIDADE FEDERAL DE CAMPINA GRANDE. **House Of Cunha**. Disponivel em: <http://houseofcunha.com.br/>. Acesso em: 18 mar. 2018. |
| 2. | LABORATÓRIO ANALYTICS - UNIVERSIDADE FEDERAL DE CAMPINA GRANDE. **Quem me representa**. Disponivel em: <http://qmrepresenta.com.br/>. Acesso em: 19 mar. 2018. |
| 3. | ROCHA, C. Site permite que eleitor veja se posições de deputados federais batem com as suas. **ESTADÃO**, São Paulo, ago. 2015. Disponivel em: <https://link.estadao.com.br/noticias/inovacao,site-permite-que-eleitor-veja-se-posicoes-de-deputados-federais-batem-com-as-suas,10000029119>. Acesso em: abr. 2018. |
| 4. | CABRAL, F. B. Como Funciona a Operação Serenata de Amor. **Medium**, 2016. Disponivel em: <https://medium.com/data-science-brigade/como-funciona-a-opera%C3%A7%C3%A3o-serenata-de-amor-25ba256e0e11>. Acesso em: 20 maio 2018. |
| 5. | OPERAÇÃO Serenada de Amor. **Operação Serenada de Amor**, 2017. Disponivel em: <https://serenata.ai/>. Acesso em: 20 maio 2018. |
| 6. | CABRAL, F. B. Operação Serenara de Amor. **Data Science Brigade**, 2017. Disponivel em: <- https://medium.com/data-science-brigade/serenata-de-amor-para-n%C3%A3o-ser-um-amor-de-ver%C3%A3o-7422c9e10fa5>. Acesso em: 22 maio 2018. |
| 7. | COLUMBUS, L. LinkedIn's Fastest-Growing Jobs Today Are In Data Science And Machine Learning. **Forbes**, 2017. Disponivel em: <https://www.forbes.com/sites/louiscolumbus/2017/12/11/linkedins-fastest-growing-jobs-today-are-in-data-science-machine-learning/#36431f9751bd>. Acesso em: 24 jun. 2018. |
| 8. | AKERKAR, R.; SAJJA, P. S. **Intelligent Techniques for Data Science**. Switzerland: Springer International Publishing Switzerland, 2016. |
| 9. | AALST, W. V. D. **Process Mining - Data Science in Action**. 2nd. ed. [S.l.]: Springer-Verlag Berlin Heidelberg, 2016. |
| 10. | MARTIN HILBERT, P. L. **The World’s Technological Capacity to Store, Communicate, and Compute Information**. [S.l.]: American Association for the Advancement of Science, v. 332 no. 6025 60-65, 2011. |
| 11. | MANYIKA, J. et al. **Big Data:** The Next Frontier for Innovation, Competition, and Productivity. [S.l.]: McKinsey, 2011. |
| 12. | DEPARTMENT OF COMPUTING AND MATHEMATICS. **Data Science and Big Data Computing**. [S.l.]: Springer International Publishing Switzerland, 2016. |
| 13. | VERBEEK, H. M. W.; BASTEN, T.; AALST, W. M. P. V. D. Diagnosing Workflow Processes using Woflan. **The Computer Journal**, v. 44, n. 4, p. 246–279, January 2001. |
| 14. | KOHLHAMMER, J. et al. Solving Problems with Visual Analytics. **Procedia Computer Science**, v. 7, p. 117-120, 2011. |
| 15. | GARRETT, J. J. **The Elements of User Experience:** User-Centered Design for the Web and Beyond. 2. ed. [S.l.]: New Riders, 2010. |
| 16. | UX - A experiência do usuário. **Think with Google**, 2015. Disponivel em: <https://www.thinkwithgoogle.com/intl/pt-br/marketing-resources/ux-e-design/ux-user-experience/>. Acesso em: 20 março 2018. |
| 17. | TSE. O TSE. **Tribunal Superior Eleitoral**. Disponivel em: <http://www.tse.jus.br/o-tse/sobre-o-tse/apresentacao>. Acesso em: 26 jun. 2018. |
| 18. | TRANSPARÊNCIA, M. D.; UNIÃO, C.-G. D. O que você encontra no Portal. **Portal da Transparência**, 2018. Disponivel em: <http://www.portaltransparencia.gov.br/sobre/o-que-voce-encontra-no-portal>. Acesso em: 27 maio 2018. |
| 19. | PETRS, J. F.; PEDRYCZ, W. **Engenharia de software – Teoria e Prática**. Rio de Janeiro: Campus, 2001. |
| 20. | O que é a computação em nuvem. **AWS**, 2018. Disponivel em: <https://aws.amazon.com/pt/what-is-cloud-computing/>. Acesso em: 20 jun. 2018. |
| 21. | GILLEANSES, G. T. A. **UML 2:** uma abordagem prática. São Paulo: Novatec, 2011. |
| 22. | ALVES, P. No GDF e no governo federal, comissionados sem vínculo são mais de 40%. **Metropoles**, 20 mar. 2018. Disponivel em: <https://www.metropoles.com/brasil/no-gdf-e-no-governo-federal-comissionados-sem-vinculo-sao-mais-de-40>. Acesso em: 30 mar. 2018. |
| 23. | FUNDAÇÃO GETULIO VARGAS. **Dapp**. Disponivel em: <http://dapp.fgv.br/ >. Acesso em: 20 mar. 2018. |
| 24. | INFOGRAM. Visualização de Dados. **Infogram**. Disponivel em: <https://infogram.com/page/visualizacao-de-dados>. Acesso em: 03 março 2018. |
| 25. | MINISTROS e secretários deixam cargos para votar sobre impeachment. **G1**, 13 abril 2016. Acesso em: 29 março 2018. |
| 26. | STANIER, C.; DEDIź, N. An Evaluation of the Challenges of Multilingualism in Data Warehouse Development, Roma, p. 196-206, abril 2016. |
| 27. | HAN, J.; KAMBER, M., 2006. Disponivel em: <http://liacs.leidenuniv.nl/~bakkerem2/dbdm2007/05\_dbdm2007\_Data%20Mining.pdf>. Acesso em: 30 março 2018. |
| 28. | MERRICK, C. 9 Reasons Data Warehouse Projects Fail. **rjmetrics**, 2014. Disponivel em: <https://blog.rjmetrics.com/2014/12/04/10-common-mistakes-when-building-a-data-warehouse/>. Acesso em: 27 março 2018. |
| 29. | EXPLORING DATA WAREHOUSES AND DATA QUALITY. **Spotless**. Disponivel em: <https://spotlessdata.com/blog/exploring-data-warehouses-and-data-quality>. Acesso em: 20 março 2018. |
| 30. | STEELE, B.; REDDY, S.; CHANDLER, J. **Algorithms for Data Science**. Switzerland: Springer International Publishing, 2016. |
| 31. | **Ferramenta utilizada para criação de mocks de tela**. Disponivel em: <https://balsamiq.com/ >. |
| 32. | PROJEÇÃO da População. **IGBE**. Disponivel em: <https://www.ibge.gov.br/apps/populacao/projecao>. Acesso em: 02 abril 2018. |
| 33. | OAB vai ao STF para limitação de cargos comissionados na esfera federal. **Contas Abertas**. Disponivel em: <http://www.contasabertas.com.br/site/orcamento/oab-vai-ao-stf-para-limitacao-de-cargos-comissionados-na-esfera-federa>. Acesso em: 02 abril 2018. |
| 34. | PIRES, A. A farra dos cargos em comissão: Brasil cabide. **Jusbrasil**, 2012. Disponivel em: <https://antoniopires.jusbrasil.com.br/artigos/121940622/a-farra-dos-cargos-em-comissao-brasil-cabide>. Acesso em: 04 abril 2018. |
| 35. | TEELE, B.; REDDY, S.; CHANDLER, J. **Algorithms for Data Science**. Switzerland: Springer International Publishing, 2016. |
| 36. | CAROSIA, A. E. D. O.; CIFERRI, C. D. D. A. Disponivel em: <http://wiki.icmc.usp.br/images/3/3c/BDAmineracao.pdf>. Acesso em: 04 abril 2018. |
| 37. | G1. **Globo**. Disponivel em: <https://g1.globo.com/politica/noticia/servidoras-do-governo-tem-melhor-formacao-mas-ocupam-menos-cargos-de-chefia.ghtml>. Acesso em: 16 mar. 2018. |
| 38. | RHYNE, T.-M. et al. Information and scientific visualization: separate but equal or happy together at last. **IEEE**, Seattle, out. 2003. Disponivel em: <https://ieeexplore.ieee.org/document/1250428/>. |

x