Cadastro Nacional de Condenações Cíveis por Ato de Improbidade Administrativa e Inelegibilidade

Relatório preliminar: descrição dos dados e inconsistências

Associação Brasileira de Jurimetria e Instituto Não Aceito Corrupção

14 de outubro de 2016

1 Introdução

O Cadastro Nacional de Condenações Cíveis por Ato de Improbidade Administrativa e Inelegibilidade (CNC) é uma base de dados controlada pelo Conselho Nacional de Justica (CNJ). A partir dessa ferramenta, é possível consultar nomes de empresas ou pessoas físicas e acessar informações sobre os processos.

O problema é que, infelizmente, o acesso à base do CNC é limitado. Com o uso de captchas e consultas individuais, não é possível obter automaticamente uma lista de todos os processos de improbidade contidos na base, o que impede o cálculo de estatísticas básicas sobre o problema e o devido acompanhamento dos processos.

O presente trabalho tem como objetivo principal solucionar este problema. A partir da construção dos programas para extração de dados, serão levantadas estatísticas básicas sobre os processos contidos no CNC, o que possibilitará maior conhecimento sobre as ações de improbidade administrativa no Brasil.

Nesta fase inicial, realizamos o download e arrumação dos dados. A metodologia para extração foi descrita próxima Seção. Todos os códigos utilizados para extração dos dados são públicos e podem ser acessados neste link.

2 Metodologia

Coletamos no total quatro bases de dados relacionais utilizando raspagem de dados (*web scraping*). Essa técnica consiste em construir um robô que acessa automaticamente diversas páginas da web e salva os dados obtidos em um computador.

2.1 Condenações

A base de condenações foi obtida em dois passos: primeiro baixamos as páginas e depois baixamos as condenações.

2.1.1 Páginas de busca

A primeira base baixada são as paginações da pesquisa. Observe este link para verificar o conteúdo de uma página. O resultado final dessa extração é uma base de dados em que cada linha é um resultado da busca indicando uma combinação de "pessoa" e "processo", e com as seguintes colunas:

- `arq`: nome do arquivo baixado, que representa uma página da pesquisa.
- `id_pag`: id da condenação (1 a 15 condenações), que repete por página.
- `id_condenacao`: id da condenação (são os números no final de `link_condenacao`).
- `id_processo`: id do processo (são os números no final de `link_processo`).
- `lab_pessoa`: nome da pessoa (ou empresa) envolvida na condenação.
- `lab_processo`: número do processo.
- `link_condenacao`: link para acesso a mais informações da condenação.
- `link_processo`: link para acesso a mais informações do processo.

A base de páginas será incorporada na base de condenações descrita a seguir.

2.1.2 Dados sobre condenações

A segunda base é obtida dos links das condenações. Acesse este link para um exemplo. Após arrumar os dados, ficamos com uma base em que cada uma das 35.977 linhas é uma **condenação** e com as seguintes 38 colunas:

- Metadados e identificadores:
  - `arq_pag`: arquivo que contém a página de pesquisa de onde foi obtido o link da condenação.
  - `id_pag`: id da condenação (1 a 15 condenações), que repete por página.
  - `arq`: arquivo que contém a página HTML com os dados da condenação.
  - `id_condenacao`: id único da condenação.
  - `id_processo`: id único do processo.
  - `id_pessoa`: id único da pessoa.
- Informações básicas:
  - `tipo_pena`: Trânsito em Julgado ou Órgão colegiado.
  - `dt_pena`: Data da pena.
  - `cod_assunto_[1:5]`: códigos dos assuntos (entre 1 e 5 assuntos) da condenação.
  - `nm_assunto_[1:5]`: nomes dos assuntos (entre 1 e 5 assuntos) da condenação.
- Inelegibilidade:
  - `teve_inelegivel`: sim ou vazio.
- Perda de Emprego/Cargo/Função Pública:
  - `teve_perda_cargo`: sim ou vazio.
- Pagamento de multa:

  - `teve_multa`: sim ou vazio.
  - `vl_multa`: valor da multa em reais.
- Ressarcimento integral do dano:
  - `teve_ressarcimento`: sim ou vazio.
  - `vl_ressarcimento`: valor da multa em reais.
- Perda de bens ou valores acrescidos ilicitamente ao patrimônio:
  - `teve_perda_bens`: sim ou vazio.
  - `vl_perda_bens`: valor da multa em reais.
- Pena privativa de liberdade:
  - `teve_pena`: sim ou vazio.
  - `duracao_pena`: duração em dias.
  - `de_pena`: data de início.
  - `ate_pena`: data do fim (pode ser no futuro).
- Suspensão dos Direitos Políticos:
  - `teve_suspensao`: sim ou vazio.
  - `duracao_suspensao`: duração em dias.
  - `de_suspensao`: data de início.
  - `ate_suspensao`: data do fim (pode ser no futuro).
- Proibição de Contratar com o Poder Público ou receber incentivos fiscais ou creditícios, direta ou indiretamente, ainda que por intermédio de pessoa jurídica da qual seja sócio majoritário:
  - `teve_proibicao`: sim ou vazio.
  - `duracao_proibicao`: duração em dias.
  - `de_proibicao`: data de início.
  - `ate_proibicao`: data do fim (pode ser no futuro).

2.2 Processos

Em seguida, obtivemos os dados de todos os processos. Após arrumar os dados, ficamos com 26.825 processos e com as seguintes 10 colunas:

- `arq_processo`: nome do arquivo (contém o id que aparece no link da base `cnc_pags`).
- `id_processo`: código identificador do processo.
- `dt_cadastro`: data de cadastro do processo no sistema.
- `n_processo`: número identificador do processo.
- `esfera_processo`: estadual, federal, militar ou superior.
- `tribunal`: nome do tribunal.
- `instancia`: primeiro grau, segundo grau, militar ou superior.

- `comarca_secao`: nome da comarca ou seção (aplicável somente ao primeiro grau).
- `vara_camara`: nome da vara (primeiro grau) ou câmara/seção de julgamento (segundo grau ou militar).
- `dt_propositura`: data de propositura da ação.

A base apresenta apenas duas inconsistências. A primeira é de um único caso que não apresenta informações em geral. A segunda são 60 casos com duas linhas cada e com números de processos idênticos na mesma instância, que podem ser cadastros duplicados.

2.3 Pessoas

Finalmente, a base de pessoas é obtida a partir de links identificados nas páginas de condenações. Cada uma das 30.541 linhas corresponde a uma pessoa (física ou jurídica), com as seguintes colunas:

- Identificadores:
  - `arq_pessoa`: nome do arquivo que contém as informações.
  - `id_pessoa`: id da pessoa (para juntar com a base de condenações).
- Informações básicas:
  - `tipo_pessoa`: F = física e J = jurídica.
  - `nm_pessoa`: Nome da pessoa.
  - `sexo`: F = feminino e M = masculino.
  - `publico`: S = funcionário público; N = não é funcionário público.
- Informações de funcionários públicos:
  - `esfera`: F = Federal, D = Distrital, E = Estadual, M = Municipal.
  - `orgao`: órgão que a pessoa trabalha (prefeitura, tribunal etc).
  - `cargo`: cargo que a pessoa exerce (prefeito, servidor etc).
  - `uf`:.
  - `cod`: código interno da pessoa (provavelmente não será utilizado).

A base apresenta algumas inconsistências. Primeiramente, temos 826 pessoas classificadas como pessoa física, mas sem informação de sexo. Além disso, temos 11 casos de pessoas classificadas como pessoa jurídica e que constam como funcionárias públicas. Dentre as 6930 pessoas classificadas como funcionárias públicas, temos 109 vazios na esfera, 343 vazios no órgão, 474 vazios no cargo e 16 vazios na UF.

2.4 Base unificada

Para facilitar as análises, construímos também uma base unificada, contendo todas as informações de condenações, pessoas e processos. Nessa base, informações sobre pessoas e processos aparecem duplicadas quando fazem parte

de mais de uma condenação. A base possui 35.977 linhas (a mesma quantidade da base de condenações) e 57 colunas.

3 Próximos passos

Com a base de dados baixada e arrumada, passaremos a realizar diversas análises no cadastro de condenados. As análises abordarão sobre valores envolvidos, pessoas condenadas e características dos processos.