

Extração e Análise de Informações Jurídicas Públicas

DCC
IME
USP

Alessandro Calò
ac@ime.usp.br

Prof. Dr. Marcelo Finger
mfinger@ime.usp.br

INTRODUÇÃO

O STF disponibiliza os textos de todos os seus processos, ou “acórdãos”, online. Tais dados podem ser lidos na página do site, mas não há um jeito de analisá-los conjuntamente. Este trabalho propõe-se a aglomerar todos os dados e fornecer estatísticas e métodos de pesquisa sobre eles. Deste modo, poder-se-á saber mais sobre o STF, analisar as suas decisões e incentivar pesquisas futuras sobre este ou qualquer outro tema.

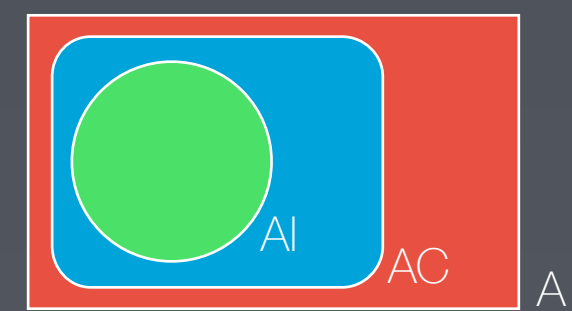
AD 157
ACO 1368
AC 012
RI 464



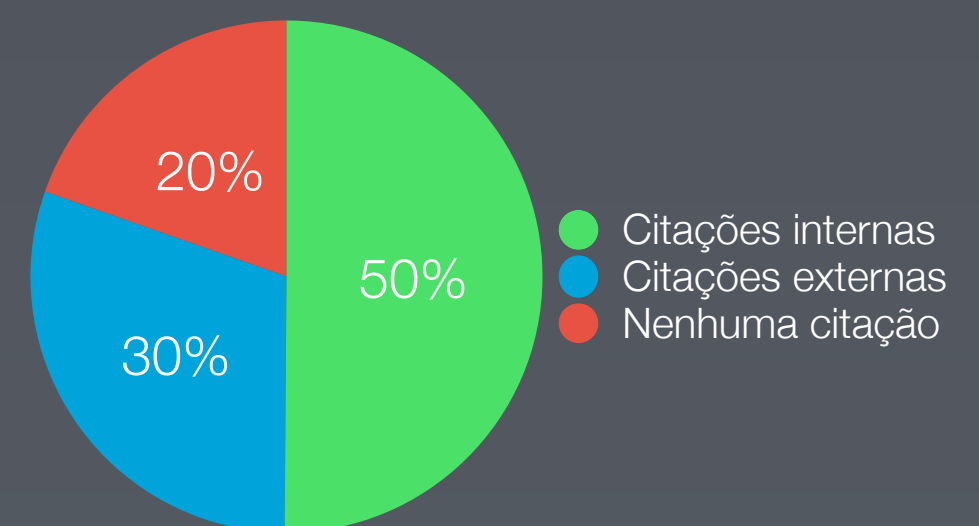
imagem: árvore de prefixos gerada a partir dos 4 identificadores acima

Seja C o número total de citações internas, detectáveis através da árvore de prefixos.

Seja A o número de acórdãos, AC o número de acórdãos que citam outros acórdãos, e AI o número de acórdãos que citam outros acórdãos da coleção.



Tais números se relacionam deste modo:



e as médias de citações encontradas são:

$$\begin{aligned}C / A &= 1.42 \\C / AC &= 1.77 \\C / AI &= 2.84\end{aligned}$$

O acórdão que mais cita é o **AP 470**, a.k.a. “mensalão”, com **176** citações

Os acórdãos mais citados são:

AI 330970 AgR - citado por **292** acórdãos
AI 372358 AgR - citado por **270** acórdãos
AI 481531 AgR - citado por **266** acórdãos

As tags mais freqüentes são:

ausência	decorrência
descabimento	existência
recurso	necessidade
extraordinário	ocorrência
voto vencido	inexistência
impossibilidade	

Os acórdãos com maior page rank são:

AI 541696 AgR	- 1.01287×10^{-3}
AI 372358 AgR	- 7.98241×10^{-4}
ADI 3289	- 7.55900×10^{-4}
ADI 1721	- 7.11430×10^{-4}
Inq 2206 QO	- 6.45409×10^{-4}

ACÓRDÃO

Um acórdão é uma decisão do órgão colegiado de um tribunal. É o relato do desfecho que um processo teve. Cada acórdão possui um nome identificador único, e características comuns como data, nome do relator, estado, etc. Além disso, cada documento pode ter uma lista de temas sobre os quais trata (a chamada “indexação”), e uma lista de acórdãos com os quais compartilha o assunto ou direcionamento jurídico (os chamados “acórdãos citados”).

EXTRAÇÃO

Para extração e download dos textos dos processos, foi utilizada a ferramenta Scrapy. Com ela foi possível dividir os processos em arquivos de texto individuais pré-processados, e armazená-los localmente. Em seguida, os arquivos foram parseados em Java, e deles foi retirada a maioria das informações.

BUSCA DE PADRÕES

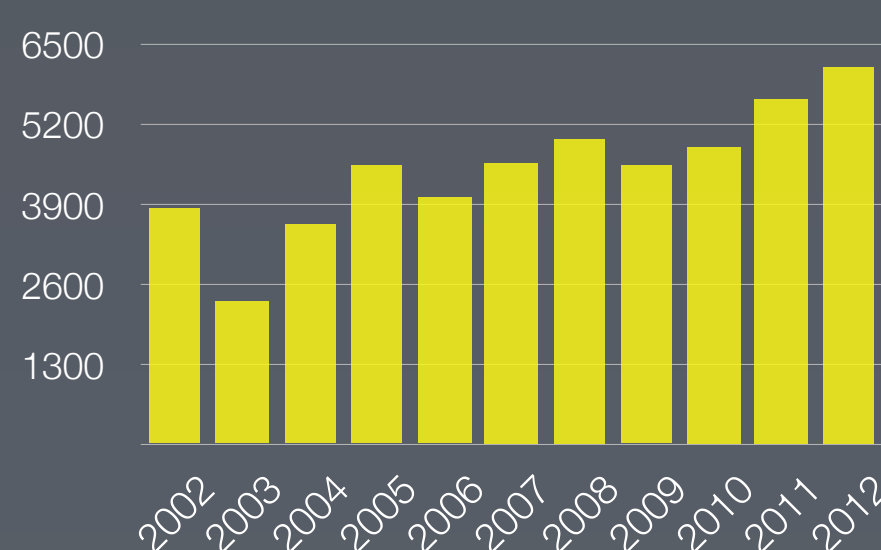
Buscar os acórdãos citados dentro do texto é uma tarefa muito difícil, pois quase não há padronização. Por isso, foi feita uma árvore de prefixos contendo os identificadores de todos os acórdãos analisados, que foi usada para procurar o identificador de cada acórdão no texto de todos os outros.

PAGE RANK

A estrutura dos acórdãos que se citam é análoga a um grafo dirigido. Foi possível então aplicar sobre estes dados o algoritmo Page Rank, desenvolvido por Larry Page e Sergey Brin em 1997, que ordena os nós de uma rede por importância, baseando-se no número e qualidade das citações feitas por um nó.

RESULTADOS

acórdãos por ano



acórdãos por estado

