Extração e Análise de Informações Jurídicas Públicas

Orientador: prof. Dr. Marcelo Finger
Orientando: Alessandro Calò

1 Motivação

Recentemente, o Supremo Tribunal Federal e o Supremo Tribunal de Justiça vêm se empenhando para tornar pública e acessível uma série de informações importantes. Trata-se da publicação em meio eletrônico dos textos de todos os acórdãos, jurisprudências e outros tipos de decisões jurídicas de última instância. A análise de tais informações pode ressaltar e ampliar a importância da publicação destes conteúdos.

Primordialmente, pode-se analisar tais dados para medir a relevância das leis e jurisprudências. Outra aplicação é o estudo de tendências, padrões e correlações no conjunto das informações disponíveis. Este conhecimento aumenta a segurança jurídica da população e pode melhorar a utilização dos serviços do judiciário.

O maior empecilho no processo de análise destes dados é o grande volume de processos disponíveis. Apesar do sofisticado sistema de buscas implementado nos sites dos tribunais, só está disponível para o usuário a análise pontual de um processo. Para o estudo global sugerido, é necessário que as informações sejam catalogadas, filtradas e condensadas em elementos demonstrativos de mais fácil manipulação e análise.

2 Objetivos

O objetivo principal desta proposta é criar um *corpus* de informação sobre jurisprudências publicadas pelos tribunais. O seu conteúdo terá informações condensadas sobre os processos analisados e permitirá a construção de uma rede de informações e dependências conceituais. Outro objetivo importante, a ser perseguido após completar o primeiro, é testar quais indicadores podem ser aplicados sobre estes dados para análises de interesse social sobre a atuação dos tribunais.

3 Métodos

3.1 Construção do *corpus*

A fase inicial consiste em obter, por extração automatizada, o conteúdo dos processos disponibilizados publicamente nos sites dos tribunais citados. Para tanto, serão utilizadas técnicas de *web crawling* para tornar o processo automático e padronizado.

3.2 Extração de informação

Após construir o *corpus* do projeto, desenvolveremos algoritmos de extração de informações textuais para obter referências que cada jurisprudência faz a outra, bem como a leis nas quais as jurisprudências se baseiam.

3.3 Algoritmo *Page Rank*

Uma das análises propostas trata de construir uma rede em que cada nó representa uma jurisprudência. Uma aresta dirigida de um nó j_1 a j_2 indica que a jurisprudência j_1 se referiu à jurisprudência j_2. Para ordenar os nós desta rede, modelada matematicamente como um grafo dirigido, pretendemos aplicar o algoritmo *Page Rank* [1], que fornece uma ordenação que pode ser analisada com a importância de um nó em relação aos demais membros da rede.

4 Resultados Esperados

4.1 *Corpus* do STF / STJ

A implementação e execução dos algoritmos citados anteriormente gerará uma base de dados sobre a qual poderemos executar os próximos passos. Além disso, esse *corpus* poderá ser útil para outras análises futuras. Para isso, serão pesquisadas e discutidas diferentes técnicas de armazenamento de dados, levando em conta a complexidade do processo e a facilidade de acesso aos dados.

4.2 Medidas de relevância

Após construir, analisar e processar o *corpus* obtido, seremos capazes de realizar medidas de relevância para determinar quais indicadores são os mais adequados para a obtenção dos dados finais, e quais são os que geram resultados mais interessantes.

Os dados deverão permitir uma visão mais clara e objetiva sobre o conjunto de jurisprudências dos dois tribunais e exibir padrões e correlações interessantes entre os acórdãos.

Referências

[1] Sergey Brin and Lawrence Page. *The Anatomy of a Large-Scale Hypertextual Web Search Engine*. 1998.