

**PLANO DE ATIVIDADES INDIVIDUAL PARA BOLSA DE INICIAÇÃO CIENTÍFICA <sup>1</sup>([fapesp.br/bolsa/ic](http://fapesp.br/bolsa/ic))**

<b>Duração da Bolsa:</b> 24 meses	<b>Dedicação:</b> 20 h semanais
-----------------------------------	---------------------------------

**1) Dados do Projeto a que se vincula a Bolsa:**

<b>Título: Centro Multidisciplinar de Estratégia, Pesquisa e Inovação da Procuradoria Geral do Estado de São Paulo - CeMEPI-PGE</b>
<b>No. Processo:</b>
<b>Pesquisador Responsável: Luciana Romano Morilas</b>
<b>Instituição/Empresa: Universidade de São Paulo</b>

**2) Dados do Responsável a quem se vincula a Bolsa:**

<b>Orientador: Rafael de Freitas Souza</b>
<b>Instituição/Empresa: Universidade de São Paulo</b>

**3) TÍTULO DO PLANO DE PESQUISA e/ou DESENVOLVIMENTO TECNOLÓGICO a ser desenvolvido pelo(a) bolsista:**

Desenvolvimento de scripts para extração e transformação automatizada de dados processuais
--

**4) RESUMO do trabalho a ser desenvolvido:**

Os dados dos processos judiciais são de livre acesso pelos websites dos tribunais de Justiça. Objetiva-se realizar a coleta utilizando-se a técnica do web scraping, que é o processo de coleta de dados estruturados da web de maneira automatizada. Outra fonte de dados será o sistemas Attornatus, cuja extração deverá ser facilitada. Além disso, os dados brutos coletados necessitarão de ajustes, limpeza e estruturação. Assim, o trabalho a ser desenvolvido consiste na implementação de scripts para a realização de web scraping do Tribunal. Além da coleta, os scripts ainda deverão organizar e estruturar os dados para serem inseridos no banco de dados, tanto dos dados coletados do Tribunal quanto do sistema Attornatus.
--

**5) OBJETIVOS pretendidos:**

<ul style="list-style-type: none"><li>- Implementação de scripts para realização coleta de dados dos processos por meio de web scraping</li><li>- Implementação de scripts para transformação e estruturação dos dados coletados a partir de todas as fontes</li><li>- Implementação de scripts para realização de inserção dos dados coletados no banco de dado</li></ul>
--

<sup>1</sup> Formulário adaptado pelo proponente a partir do formulário de bolsas TT

**6) PLANO DE TRABALHO (inclua os procedimentos e métodos, ferramentas e técnicas, além das atividades a serem desempenhadas pelo bolsista e a Tabela com Cronograma de Execução das atividades).**

O desenvolvimento será pautado pelo ciclo de vida de desenvolvimento de sistemas, de acordo com o que é preconizado na área de engenharia de software, a saber:

1. Análise: levantamento dos requisitos para o desenvolvimento. Estudos sobre técnicas de web scraping. Análise da estrutura dos websites dos tribunais.
2. Projeto: elaboração do fluxo de execução dos scripts, com possível determinação de módulos e relacionamentos entre eles.
3. Implementação: codificação dos scripts
4. Testes: realização de testes para verificação da assertividade da coleta. Realização de ajustes e correções.
5. Produção: execução dos scripts, realizando a coleta de dados.
6. Escrita de relatórios

Ativ	Mês 1	Mês 2	Mês 3	Mês 4	Mês 5	Mês 6	Mês 7	Mês 8	Mês 9	Mês 10	Mês 11	Mês 12
1												
2												
3												
4												
5												
6												

O acompanhamento da execução das atividades será semanal, sendo gerenciado por metodologia ágil. Os códigos serão gerenciados pela plataforma GitHub que auxilia na hospedagem de código-fonte e arquivos com controle de versão usando o Git.

Segue o cronograma das etapas supracitadas (essa bolsa será implementada no primeiro ano do projeto)