
BOLSA DE ESTUDOS – PROCESSO SELETIVO

O Escritório de Apoio à Pesquisa comunica a abertura de processo seletivo para indicação de 1 bolsista na modalidade iniciação científica, para atuar sob orientação do Prof. Dr. Evandro Marcos Saidel Riberiro no âmbito do projeto “Modelagem e implementação de uma base de dados aberta para fins de análise de decisões judiciais” (ANEXO I), vinculado ao grupo de pesquisas Applied Business Law.

1. REQUISITOS

Vagas exclusivas para alunos regularmente matriculados nos cursos de graduação da USP, cumprindo os seguintes requisitos:

Estar cursando Ciência da Computação, Informática Biomédica ou outros cursos afins (a partir do 3º. ano/6º. Semestre), com conhecimentos em:

- Sistema de gerenciamento de banco de dados (desejado, mas não restrito: NoSQL, mais especificamente MongoDB).
- Linguagem de programação Python (desejado, mas não restrito: conhecimento em web scrapping)

2. VALOR E VIGÊNCIA

Valor: R\$ 450,00 mensais para a carga horária de 20 horas semanais.

Vigência: de 01/04/2021 a 31/03/2022.

3. ATIVIDADES

Vide ANEXO I

4. INSCRIÇÃO

As inscrições serão efetuadas através de mensagem eletrônica enviada ao endereço apoiopq@fearp.usp.br, com o título INSCRIÇÃO – PROCESSO SELETIVO BOLSA – ABL2021, no período de 24/03 a 31/03/2021, e deverá conter os seguintes documentos anexados:

- Carta de motivação indicando como o candidato atende aos requisitos necessários;
- Histórico Escolar de graduação;

5. SELEÇÃO

A seleção será realizada pela análise dos documentos listados no item 4 e eventual entrevista presencial com o coordenador do projeto.

6. RESULTADO

O resultado será divulgado a partir em 01/04/2021 na página eletrônica da Pesquisa no sítio da FEA-RP (<https://www.fearp.usp.br/pt-br/pesquisa>). O(a) candidato(a) selecionado(a) será contatado por email em 01/04/2021, para início imediato das atividades.

ANEXO 1

Identificação do Projeto

Título: Modelagem e implementação de uma base de dados aberta para fins de análise de decisões judiciais

Proponente: Prof. Dr. Evandro Marcos Saidel Ribeiro

Palavras-chave: Data science, Banco de dados NoSQL, Jurimetria, Tomada de decisão

1. Contextualização

Atualmente, a informação tem sido reconhecida como um ativo estratégico para qualquer organização e área de conhecimento. A gestão e a liderança eficazes envolvem, principalmente, a solução criativa de problemas, motivando os funcionários e garantindo que a organização, seja ela pública ou privada, atinja seus objetivos e metas.

Os dados obtidos de fontes internas e externas se constituem nesse ativo estratégico e são um desafio particularmente grande para os modernos sistemas de Tecnologia da Informação (TI), porque tornam esses sistemas muito lentos e não escaláveis o suficiente (LAUDON; LAUDON, 2015).

No que concerne a área de Direito, o acesso à informação é a ferramenta mais fundamental para garantir acesso qualificado à justiça. Apesar da previsão constitucional existente desde 1988 e que determina o acesso à informação como um direito individual dos cidadãos, o regramento infraconstitucional começou a regulamentar essas questões tardiamente.

Por um lado, aproveitando o poder dos recursos computacionais, os advogados podem prever com mais precisão como os eventos se desenrolam nos litígios. Assim, os advogados que adotarem a tomada de decisão baseada em dados terão uma clara vantagem sobre seus colegas que ainda se apegam a seus instrumentos desatualizados. (LETTIERI et al., 2018)

Por outro lado, o Conselho Nacional de Justiça (CNJ), os Tribunais e os magistrados, tem a possibilidade, com a utilização massiva de recursos e técnicas de TI, de analisar cenários e traçar estratégias de ação a fim de minimizar, ou mesmo solucionar, notórios problemas como, por exemplo, a demora na prestação jurisdicional e a consequente ineficiência do judiciário.

No cenário das políticas judiciárias, a evolução qualitativa e quantitativa dos serviços oferecidos à sociedade envolve a identificação das principais dificuldades e oportunidades existentes em relação à formulação, à implementação e ao controle das atividades de prestação jurisdicional na busca de concretização de direitos e garantias fundamentais (CNJ, [s.d.]).

Neste contexto atua o grupo de pesquisas Applied Business Law da FEA-RP. Por meio de aplicações de técnicas de jurimetria, uma equipe multidisciplinar analisa vários temas e assuntos pertinentes a Justiça no escopo de negócios, visando indicar melhores cenários de tomada da decisão tanto para advogados, quanto magistrados e mesmo organizações.

Assim, este projeto de iniciação científica tem como objetivo a modelagem e implementação da base de dados utilizada pelo grupo de pesquisa. Mais especificamente o bolsista será responsável pela atualização da base de dados já existente e também por implementações de melhorias.

2. Metodologia

2.1 Descrição e abrangência do universo de pesquisa

O universo da pesquisa consistirá em coleta dos dados disponíveis no contexto digital disponíveis por meio das decisões publicadas nos respectivos diários oficiais e/ou sites dos Tribunais, de acordo com as necessidades e problemas de pesquisa do grupo Applied Business Law.

2.2 Detalhamento da metodologia de coleta e geração de dados

A mineração de dados é uma tecnologia poderosa que pode ser definida como o processo automatizado de extração de conhecimento e informações úteis, incluindo padrões, associações, ou

seja, é uma tecnologia que combina métodos tradicionais de análise de dados com algoritmos sofisticados para o processamento de grandes volumes de dados (VENKATADRI; REDDY, 2011).

Han; Kamber; Pei (2012) descrevem o processo de descoberta de conhecimento como uma sequência iterativa das seguintes etapas:

1. Limpeza de dados: para remover o ruído e dados inconsistentes;
2. Integração de dados: etapa em que várias fontes de dados podem ser combinadas;
3. Seleção de dados: os dados relevantes para a tarefa de análise são recuperados do banco de dados;
4. Transformação: os dados são transformados e consolidados em um padrão apropriado para mineração, realizando operações de resumo ou agregação;
5. *Data mining*: processo essencial em que métodos são aplicados para extrair padrões de dados;
6. Avaliação de padrões: para identificar os padrões verdadeiramente interessantes que representam o conhecimento com base em medidas de interesse, e;
7. Apresentação: etapa em que a visualização e as técnicas de representação do conhecimento são usadas para apresentar conhecimento.

No que se refere ao contexto do projeto, o processo de descoberta do conhecimento anteriormente descrito será aplicado conforme esquema da Figura 1.

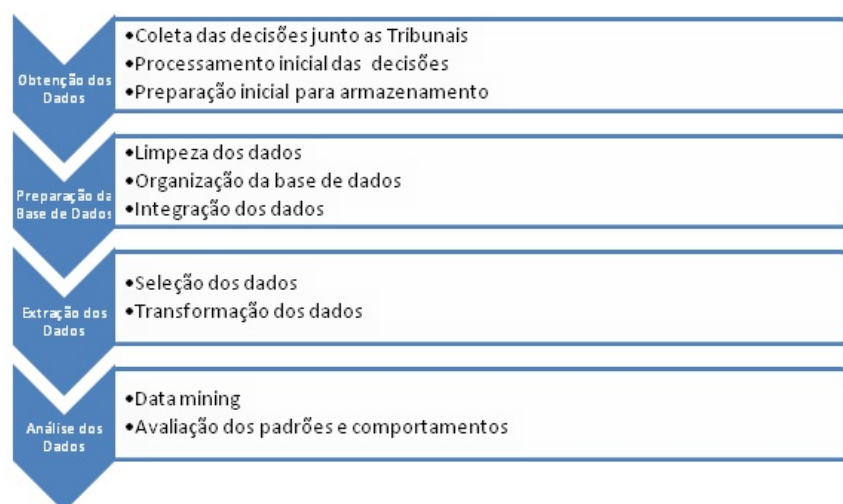


Figura 1 – Etapas para a realização da pesquisa

A etapa de obtenção dos dados acontece antes da realização da limpeza dos dados (primeira etapa do processo de descoberta do conhecimento). Para que haja sucesso em qualquer pesquisa empírica é necessário o acesso a dados confiáveis, em quantidade suficiente para que se minimizem os efeitos de vieses cuja ocorrência pode inviabilizar o resultado final desejado. Nesse aspecto, o presente projeto visa trabalhar com fontes de dados oriundas de informações oficiais públicas, por meio da captura dos mesmos dados de modo programático via sistemas informatizados, como por exemplo, web scrapping.

Tendo em vista que atualmente todos os tribunais, estaduais e federais, dispõem de sistemas informatizados de consulta aos andamentos processuais, bem como possuem um sistema de publicações oficiais por meio de diários digitais, é possível acessar uma grande base de dados que servirá de matéria prima da qual a pesquisa se alimentará.

Após a recuperação dos dados, é necessária a realização de um processamento inicial que consiste em converter tais dados em texto puro (extensão de arquivo .txt), na separação das páginas e a organização e padronização dos dados, com a finalidade de preparação inicial para o armazenamento.

Armazenados os arquivos, pretende-se organizá-los em algum sistema de gerenciamento de banco de dados (SGBD) noSQL. Em contraposição aos SGBD relacionais, os noSQL são projetados para armazenamento e processamento de dados em grande escala, principalmente para lidar com

Escritório de Apoio à Pesquisa

enormes quantidades de dados as quais soluções SGBDs convencionais não conseguiram lidar. Eles podem suportar várias atividades, incluindo análises exploratórias e preditivas, transformação de dados em no modelo ETL, gerenciando transações de longa duração ou interorganizações, por exemplo (MONIRUZZAMAN; HOSSAIN, 2013).

Depois de organizados tais dados é possível identificar os processos que se enquadrem no tema proposto filtrando todas as publicações que a eles se relacionem (extração dos dados). A partir das publicações filtradas é possível analisar quais dessas tratam-se de sentenças judiciais de primeira instância, partindo então para a análise do conteúdo desse material, identificando padrões que se alinhem com as hipóteses levantadas. É possível ainda a análise complementar do trâmite de tais processos nas varas originais por meio da captura eletrônica de seus andamentos, o que pode complementar conclusões tais como tempo efetivo de tramitação, assim como o efeito dos atos não publicáveis no deslinde de cada causa, o que possibilitaria, hipoteticamente uma análise do trabalho dos órgãos judiciais envolvidos (análise dos dados).

3. Atividades

Para a realização do projeto, pretende-se executar as seguintes atividades, obedecendo o cronograma no Quadro 1. Também será observada a obrigatoriedade dos relatórios parcial e final.

1. Reconhecimento da base atual no Servidor e das fontes de coletas de dados

A base já existente está hospedada em um servidor da Internuvem USP. A tarefa inicial é configurar reconhecer os recursos disponíveis no servidor. Também será necessário o reconhecimento do SGBD MongoDB, bem como as ferramentas acessórias. Além disso, esta etapa também prevê o conhecimento das principais fontes de dados (sites dos Tribunais e diários oficiais)

2. Atualização e Organização da base de dados

Tendo em vista a base já existente, o início da coleta dos dados será no sentido da atualização da base, com a inserção de dados coletados conforme solicitação da equipe de pesquisa. Oportunamente será necessário realizar a modelagem e estruturação da base de armazenamento com a inserção dos resultados obtidos. Como mencionado, pretende-se organizá-los em algum sistema de gerenciamento de banco de dados (SGBD) *noSQL* que são projetados para armazenamento e processamento de dados em grande escala. O sistema de gerenciamento escolhido foi o MongoDB.

3. Integração dos dados

Como explicado, os dados provem das publicações e sites dos Tribunais de Justiça dos Estados que compõem o projeto. Os dados não estão padronizados e organizados para armazenamento. Haverá a necessidade do entendimento dos campos e atributos pertinentes as pesquisas para, conseqüentemente, a estruturar a base.

4. Seleção dos dados

Uma vez estruturada e atualizada, a base de dados estará pronta para a realização de consultas para obtenção dos resultados. Nessa fase serão elaboradas as *queries* de acordo com as necessidades da equipe de pesquisa.

5. Monitoramento, melhorias e suporte para obtenção dos dados

Esse item consiste no contínuo monitoramento de melhorias, contribuindo para o desenvolvimento de consultas padronizadas por meio de uma interface amigável, facilitando a busca de informações.

Escritório de Apoio à Pesquisa

Quadro 1 - Cronograma previsto

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												

4. Resultados Esperados

O principal resultado esperado é fornecer o suporte para a realização das pesquisas de interesse dos membros do grupo de pesquisas Applied Business Law da FEA-RP. A infraestrutura lógica de armazenamento implementada será de fundamental importância para alcançar os objetivos da pesquisa.

Além disso, a base de dados ficará disponível para outros pesquisadores que queiram extrair dados para realização de suas pesquisas. Adicionalmente, o trabalho pode contribuir para o acesso à informação.

Referências

- CNJ. **CNJ Pesquisa**. Disponível em: <<http://www.cnj.jus.br/corregedoriacnj/190-gestao-planejamento-e-pesquisa/18884-cnj-pesquisa>>. Acesso em: 15 maio. 2018.
- EDITAL 02/2017. **Edital de convocação pública e de seleção convocação**. Disponível em: <<http://www.cnj.jus.br/files/conteudo/arquivo/2017/09/780a1047c628a90da5a29928f0d4dc01.pdf>>. Acesso em: 15 maio. 2018.
- HAN, J.; KAMBER, M.; PEI, J. **Data mining concepts and techniques, third edition**. 3rd. ed. Waltham, Mass.: Morgan Kaufmann Publishers, 2012.
- LAUDON, K.; LAUDON, J. **Sistemas de informações gerenciais**. 11a. ed. [s.l.] Pearson/Prentice Hall, 2015.
- LETTIERI, N. et al. Ex Machina: Analytical platforms, Law and the Challenges of Computational Legal Science. **Future Internet**, v. 10, n. 5, 26 abr. 2018.
- MONIRUZZAMAN, A. B. M.; HOSSAIN, S. A. Nosql database: New era of databases for big data analytics-classification, characteristics and comparison. **arXiv preprint arXiv:1307.0191**, v. 6, n. 4, p. 1–14, 2013.
- VENKATADRI, M.; REDDY, D. L. C. Article: A Review on Data mining from Past to the Future. **International Journal of Computer Applications**, v. 15, n. 7, p. 19–22, 2011.