

# Jusbrasil e os Desafios Tecnológicos para Facilitar e Aprimorar o Acesso à Justiça

Edleno Silva de Moura<sup>1</sup>, Rafael Costa<sup>2</sup>, Gabriel Jordão<sup>1</sup>, Gustavo Barreto Maia<sup>2</sup>

<sup>1</sup>Instituto de Computação – Universidade Federal do Amazonas (UFAM)  
Av. General Rodrigo Octavio Jordão Ramos, 1200 – Manaus – Brasil

<sup>2</sup>Jusbrasil  
Av. Tancredo Neves, 1186 – Salvador – Brasil

edleno@icomp.ufam.edu.br, {rafael,gabriel,gustavo}@jusbrasil.com.br

**Abstract.** *This article presents an introduction to some of the technological challenges faced by the company Jusbrasil in its mission to approximate the Brazilians to the justice. Jusbrasil seeks to combine law and technology so that justice crosses the borders of the courts and reaches the homes of any citizen. Millions of people access the company's platform nowadays, with access to more than 900,000 lawyers. On the other hand, our database has billions of documents containing artifacts related to law in Brazil. This scenario provides opportunities for the development of intelligent products, such as effective and efficient search systems, tools for structuring and processing information, data mining and recommendation systems, among others. In this work, we discuss and present to the academic community some of the technological challenges faced by the company.*

**Resumo.** *Este artigo apresenta uma breve introdução a alguns dos desafios tecnológicos enfrentados na empresa Jusbrasil em sua missão de facilitar o acesso à justiça no Brasil. A Jusbrasil busca unir direito e tecnologia para que a justiça ultrapasse as fronteiras dos tribunais e chegue às casas de qualquer cidadão ou cidadã. Milhões de pessoas acessam a plataforma da empresa atualmente, havendo acesso de mais de 900 mil advogados. Por outro lado, nossa base de informação possui bilhões de documentos contendo artefatos relacionados ao direito no Brasil. Essa base de informação traz oportunidades para o desenvolvimento de produtos inteligentes, tais como sistemas de busca eficazes e eficientes, ferramentas de estruturação e tratamento da informação, mineração de dados e sistemas de recomendação, dentre outros. Neste trabalho, discutimos e apresentamos para a comunidade acadêmica alguns dos desafios tecnológicos enfrentados pela empresa no dia-a-dia.*

## 1. Introdução

Para muitos a palavra justiça se resume aos tribunais, mas ela deveria fazer parte das escolhas diárias de qualquer pessoa. No meio do caminho para uma sociedade mais justa, milhões de brasileiros se deparam com um abismo entre ter um direito e torná-lo parte de suas vidas.

A Jusbrasil surgiu a partir da percepção desse problema e tem buscado unir direito e tecnologia para que a justiça ultrapasse as fronteiras dos tribunais e chegue às casas

de qualquer cidadão ou cidadã. Hoje em dia no Brasil milhões de pessoas já acessam a plataforma da empresa diariamente e mais de 900 mil advogados utilizam produtos desenvolvidos pela Jusbrasil para aumentar sua produtividade.

A empresa tem investido maciçamente em tecnologias para coleta e organização da informação jurídica. Como resultado, possui hoje uma base com bilhões de documentos contendo artefatos relacionados ao direito no Brasil, incluindo processos, jurisprudência, artigos, leis, modelos e peças jurídicas e informação publicada em diários oficiais, dentre outros. A empresa reúne assim uma das maiores bases de dados sobre a Justiça Brasileira.

Em sua jornada para entregar valor a partir da sua larga base de usuários, aliada à enorme base de dados construída, a empresa tem se deparado com diversos desafios de pesquisa interessantes. Este artigo tem como objetivo apresentar alguns desses desafios e fazer uma breve introdução ao que se tem feito na literatura e na empresa para resolvê-los. Esperamos com isso motivar mais pessoas a nos ajudar na empreitada de contribuir para levar a justiça cada vez mais perto do cidadão comum. Os desafios escolhidos para serem discutidos aqui serão mais relacionados a sistemas de busca, recomendação e estruturação automática da informação. Contudo, há muitas outras possibilidades que podem ser levantadas

## 2. Arquitetura

Antes de mostrar os desafios específicos de pesquisa que encontramos no dia-a-dia na Jusbrasil, vamos explicar em linhas gerais como funciona o fluxo de tratamento de dados dentro da empresa. A Figura 1 apresenta um diagrama que descreve como a informação flui pela empresa.

Coletores, mostrados no canto superior esquerdo da figura, capturam informações provenientes de centenas de fontes disponíveis online. Os coletores são responsáveis por alimentar as bases de dados da empresa com informação dos mais diversos artefatos jurídicos, tais como leis, diários oficiais, processos, decisões jurídicas, dentre outras. Os dados coletados são em geral capturados na forma de documentos não-estruturados, disponibilizados nos mais diversos formatos, tais como PDF, texto ou HTML.

Os documentos coletados passam então por um processo de extração e limpeza de dados, onde são convertidos para textos e também são extraídas algumas informações estruturadas. Nessa fase, busca-se, por exemplo, a identificação de entidades importantes, tais como citação a pessoas ou empresas, tribunais, etc.

Os dados passam então pelo que chamamos na figura de integração. Nessa fase são realizadas diversas tarefas que visam transformar os dados coletados em uma base de conhecimento confiável. Incluem-se nessa fase processos de identificação de citação de um documento a outros documentos já armazenados na base, a associação entre nomes extraídos e nomes contidos em nossas bases de dados, a deduplicação de dados, processos de agrupamento e classificação da informação para uso posterior, dentre outros. O processo gera como resultado o que chamamos de base de conhecimento (do inglês *knowledge base*), contendo informação tratada e estruturada que fica então disponível para o desenvolvimento de serviços para nossos usuários.

Os dados gerados são então disponibilizados em filas que alimentam sistemas de

indexação de informação que geram índices para acesso rápido em nossas aplicações, como mostrado no canto inferior direito da figura, no processo descrito como Modelagem e Indexação.

De forma complementar, a informação também entra na empresa por meio das interações dos usuários com o sistema. Nossos usuários são cadastrados e apresentam em seu cadastro informações que são usadas para melhorar nossa prestação de serviço. Da mesma forma, também usamos as interações entre os usuários, sempre com a preocupação de respeitar a privacidade de todos. Um problema a ser enfrentado com os de interação entre usuários e sistemas é que eles podem trazer informação tendenciosa. Para contornar esse problema, os dados passam por um processo de eliminação de *bias*, ou ainda eliminação de enviesamento tendencioso, antes de serem usados. A eliminação de *bias* é por si só um processo que pode trazer diversos desafios de pesquisa [Krishnan et al. 2014, Collins et al. 2018, Abdollahpouri et al. 2019].

Das interações e do cadastro, pode-se produzir informação útil para alimentar as aplicações e serviços oferecidos pela empresa. Por exemplo, podemos citar os cliques em documentos apresentados nos sistemas de busca e recomendação, a possibilidade de agrupar usuários por interesse em comum, região demográfica, etc.

De posse de tais dados, desenvolvemos então nossas aplicações que chegam até o usuário, tais como o sistema de buscas do Jusbrasil, a recomendação de conteúdo para nossos usuários, um sistema de automação de tarefas jurídicas conhecido como escritório online, etc. Outras aplicações que requerem uso intensivo de aprendizagem de máquina, classificação, mineração e estruturação de dados estão em nossos planos para o futuro.

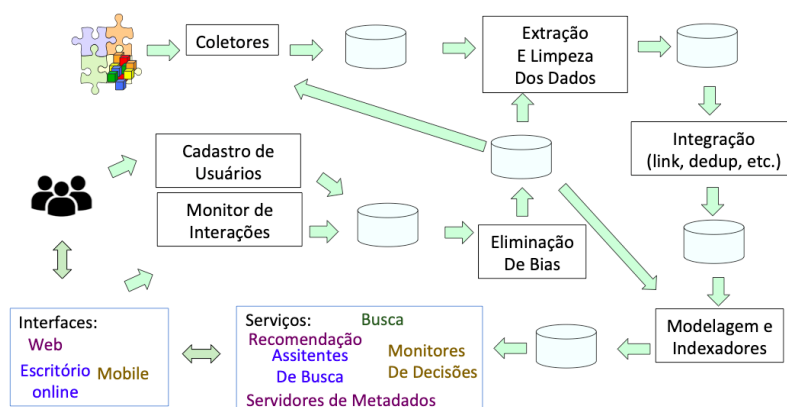


Figura 1. Visão Geral da Arquitetura dos Principais Sistemas da Jusbrasil

### 3. Alguns Desafios de Pesquisa

Nesta seção apresentamos alguns dos diversos desafios de pesquisa que podem ser encontrados no dia-a-dia da empresa. Dentre eles, citamos aqui o desenvolvimento de sistemas de busca eficazes e eficientes, o desenvolvimento de sistemas de recomendação e a produção de uma base de conhecimento confiável a partir de dados estruturados

#### 3.1. Desenvolvimento de Sistemas de Busca Eficazes e Eficientes

A empresa disponibiliza a seus usuários um sistema de busca por informação em uma base de dados contendo bilhões de documentos. Fornecer resultados precisos é uma tarefa

de vital importância e por isso perseguimos continuamente novos algoritmos e métodos para processamento de consultas eficazes e eficientes. Entre as necessidades voltadas à eficácia na entrega de resultados relevantes, são importantes estudos voltados ao uso de aprendizagem de máquina em sistemas de busca [Silva et al. 2009, Sousa et al. 2019, Guo et al. 2019]. A empresa tem investido nessa frente, bem como em infraestrutura para a coleta e disponibilização de informação que possa ser usada como insumo para que os algoritmos de aprendizagem de máquina possam gerar ordenação de resposta de alta qualidade, tais como dados de navegação, agrupamentos de usuários e informação demográfica, dentre outros. O simples estudo da melhor forma de usar tais informações abre várias oportunidades de pesquisa.

Na linha de processamento eficiente de informação o desafio também é muito grande. Apesar de busca ser tratada em geral no meio empresarial como uma *commodity*, o desenvolvimento de algoritmos eficientes de processamento de consultas pode abrir espaço não só para uma grande redução nos custos de operação da empresa, mas também garantir sua escalabilidade e a oferta de uma busca com mais qualidade e velocidade. Além disso, sistemas de busca disponíveis no mercado levam muito tempo para incorporar melhorias propostas na literatura, em parte porque algumas delas podem não beneficiar a todos os tipos de empresas que usam tais sistemas ou ainda podem até prejudicar parte de seus usuários. Exemplos de melhorias que podem impactar muito sistemas de busca como o da Jusbrasil, mesmo que não impactem tanto a outros tipos de sistemas de busca incluem novos métodos de poda e processamento de consultas [Rossi et al. 2013, Daoud et al. 2017, Petri et al. 2019] e novas técnicas para o desenvolvimento de sistemas de busca eficientes que usem aprendizagem de máquina [Lucchese et al. 2018, Silva et al. 2020].

### **3.2. Sistemas de Recomendação**

Sistemas de recomendação têm potencial para desempenhar um papel central na empresa para o futuro, podendo ser usados para recomendar artefatos jurídicos quando o usuário está visitando outro artefato jurídico em nosso site, na recomendação de profissionais para assessorar pessoas quando buscam a justiça ou no auxílio a advogados durante a produção de peças, petições, procurações e demais tipos de documentos que precisem ser gerados no dia-a-dia.

O uso de nossos sistemas por milhões de pessoas, aliado à enorme quantidade de informação existente, faz do Jusbrasil um excelente laboratório para a experimentação de novos algoritmos e métodos ligados à área de recomendação. Exemplos de desafios interessantes nessa área podem incluir o desenvolvimento de métodos de recomendação que produzam um bom equilíbrio entre novidade, diversidade e popularidade de conteúdo [Ribeiro et al. 2014], estudos de formas de explicar o resultado de recomendações [Nóbrega and Marinho 2019]. De fato, sistemas automáticos para auxiliar profissionais da área do direito podem revolucionar o trabalho na área [Kluttz and Mulligan 2019], e parte dessa revolução pode vir do desenvolvimento de métodos e estudos na área de recomendação específicos para a área.

### **3.3. Construção de uma Base de Conhecimento**

A construção de uma base de conhecimento confiável é um dos grandes desafios da empresa. Precisamos, por exemplo, extrair com precisão informação sobre valores en-

volvidos em um processo, definição clara de quem é o demandante e o demandado, agrupar processos por similaridade para ajudar na identificação de chances de sucesso para ambas as partes, dentre muitas outras tarefas que ainda precisam ser realizadas com precisão no futuro. Para chegar a ter uma base de conhecimento sólida, enfrentamos problemas como a possibilidade de pessoas distintas terem o mesmo nome, de uma mesma pessoa ou empresa ter seu nome citado de maneira ligeiramente diferente de um processo para outro, a ocorrência de citações a uma jurisprudência em diversas formas distintas, dentre outras. Surgem aí oportunidades em diversas linhas de pesquisa, tais como a de deduplicação de dados [De Freitas et al. 2010], extração e estruturação de dados a partir de texto [Cortez et al. 2011], processos de agrupamento e classificação de documentos para uso posterior [Cristo et al. 2003], dentre outros. O processo gera como resultado o que chamamos de base de conhecimento (do inglês *knowledge base*), contendo informação tratada e estruturada que fica então disponível para o desenvolvimento de serviços para nossos usuários. Surgem aí oportunidades também para o uso da informação produzida, tais como interfaces mais amigáveis para busca de informação estruturada baseada em palavras-chave [Mesquita et al. 2007] ou o desenvolvimento de sistemas mais sofisticados para predição de decisões usando nossa base de conhecimento [Zhong et al. 2018, Li et al. 2019] ou outras tarefas que visem aprimorar o trabalho da justiça por meio de automação [Li et al. 2019]

### 3.4. Outros desafios

Além dos dois exemplos que apresentamos rapidamente aqui, há muitos outros desafios de pesquisa na empresa que estão relacionados ao bom funcionamento de nosso sistema, incluindo o desenvolvimento de métodos automáticos de coleta de dados [Vidal et al. 2008], de métodos de processamento de linguagem natural voltados para a área do direito, de métodos desambiguação de dados, remoção de dados duplicados, estruturação automática da informação, dentre outros.

## 4. Conclusões

Como mostrado neste artigo, a empresa Jusbrasil tem como principal objetivo o desenvolvimento de tecnologias para aproximar pessoas da justiça, visando contribuir para que os processos envolvendo a justiça se tornem mais claros, mais eficientes e disponíveis às pessoas. A busca desse objetivo traz consigo diversos desafios tecnológicos onde a pesquisa na área de ciência da computação pode contribuir significativamente na produção de sistemas cada vez mais eficazes e efetivos.

## Referências

- Abdollahpouri, H., Burke, R., and Mobasher, B. (2019). Managing popularity bias in recommender systems with personalized re-ranking. *arXiv preprint arXiv:1901.07555*.
- Collins, A., Tkaczyk, D., Aizawa, A., and Beel, J. (2018). Position bias in recommender systems for digital libraries. In *International Conference on Information*, pages 335–344. Springer.
- Cortez, E., Oliveira, D., da Silva, A. S., de Moura, E. S., and Laender, A. H. (2011). Joint unsupervised structure discovery and information extraction. In *Proceedings of the 2011 ACM SIGMOD*, pages 541–552.

- Cristo, M., Calado, P., de Moura, E. S., Ziviani, N., and Ribeiro-Neto, B. (2003). Link information as a similarity measure in web classification. In *International Symposium on String Processing and Information Retrieval. SPIRE*, pages 43–55. Springer.
- Daoud, C. M., de Moura, E. S., Fernandes, D., da Silva, A. S., Rossi, C., and Carvalho, A. (2017). Waves: a fast multi-tier top-k query processing algorithm. *Information Retrieval Journal*, 20(3):292–316.
- De Freitas, J., Pappa, G. L., da Silva, A. S., Gonc, M. A., Moura, E., Veloso, A., Laender, A. H., de Carvalho, M. G., et al. (2010). Active learning genetic programming for record deduplication. In *IEEE Congress on Evolutionary Computation*, pages 1–8.
- Guo, J., Fan, Y., Pang, L., Yang, L., Ai, Q., Zamani, H., Wu, C., Croft, W. B., and Cheng, X. (2019). A deep look into neural ranking models for information retrieval. *arXiv preprint arXiv:1903.06902*.
- Kluttz, D. N. and Mulligan, D. K. (2019). Automated decision support technologies and the legal profession. *Berkeley Tech. LJ*, 34:853.
- Krishnan, S., Patel, J., Franklin, M. J., and Goldberg, K. (2014). A methodology for learning, analyzing, and mitigating social influence bias in recommender systems. In *Proceedings of the 8th ACM Conference on Recommender systems*, pages 137–144.
- Li, S., Zhang, H., Ye, L., Guo, X., and Fang, B. (2019). Mann: A multichannel attentive neural network for legal judgment prediction. *IEEE Access*, 7:151144–151155.
- Lucchese, C., Nardini, F. M., Perego, R., Orlando, S., and Trani, S. (2018). Selective gradient boosting for effective learning to rank. In *Proceedings of 41st SIGIR*, pages 155–164. ACM.
- Mesquita, F., da Silva, A. S., de Moura, E. S., Calado, P., and Laender, A. H. (2007). Labrador: Efficiently publishing relational databases on the web by using keyword-based query interfaces. *Information processing & management*, 43(4):983–1004.
- Nóbrega, C. and Marinho, L. (2019). Towards explaining recommendations through local surrogate models. In *Proceedings of the 34th ACM SIGAPP*, pages 1671–1678.
- Petri, M., Moffat, A., Mackenzie, J., Culpepper, J. S., and Beck, D. (2019). Accelerated query processing via similarity score prediction. In *Proceedings of the 42nd SIGIR*, pages 485–494.
- Ribeiro, M. T., Ziviani, N., Moura, E. S. D., Hata, I., Lacerda, A., and Veloso, A. (2014). Multiobjective pareto-efficient approaches for recommender systems. *ACM TIST*, 5(4):1–20.
- Rossi, C., de Moura, E. S., Carvalho, A. L., and da Silva, A. S. (2013). Fast document-at-a-time query processing using two-tier indexes. In *Proceedings of the 36th SIGIR*, pages 183–192.
- Silva, S. D. N., De Moura, E. S., Calado, P. P., and Da Silva, A. S. (2020). Effective lightweight learning-to-rank method using unified term impacts. *IEEE Access*, 8:70420–70437.
- Silva, T. P. C., de Moura, E. S., Cavalcanti, J. M. B., da Silva, A. S., de Carvalho, M. G., and Gonçalves, M. A. (2009). An evolutionary approach for combining different sources of evidence in search engines. *Information Systems*, 34(2):276–289.

- Sousa, D. X., Canuto, S., Gonçalves, M. A., Rosa, T. C., and Martins, W. S. (2019). Risk-sensitive learning to rank with evolutionary multi-objective feature selection. *ACM TOIS*, 37(2):24:1–24:34.
- Vidal, M. L., da Silva, A. S., de Moura, E. S., and Cavalcanti, J. M. (2008). Structure-based crawling in the hidden web. *J. UCS*, 14(11):1857–1876.
- Zhong, H., Guo, Z., Tu, C., Xiao, C., Liu, Z., and Sun, M. (2018). Legal judgment prediction via topological learning. In *Proceedings of EMNLP*, pages 3540–3549.