

relatorio

Luísa

05/12/2021

Relatório Final Diário do Clima

Equipe OKBR:

- Andre Assumpção;
- Ariane Alves;
- Fernanda Campagnucci;
- Giulio Carvalho;
- Juliana Trevine;
- Luísa Coelho;
- Murilo Machado;
- Pedro Guimarães

Resumo

Proposta de relatório final do projeto Diário do Clima.

Introdução

O Diário Oficial da União (DOU) é um importante veículo de comunicação nacional, criado em 1º de outubro de 1862, sancionado pela Lei Imperial nº 1.177, após 54 anos da chegada da corte portuguesa ao Brasil. Sua função, desde então, é divulgar as leis e normas do país, de forma simples, direta e documental, sem qualquer tipo de contestação ou queixa, sendo este o instrumento em que o governo presta contas à sociedade (Rocha, 2020). Atualmente o DOU pode ser acessado através do link da Imprensa Nacional permitindo buscas por palavras-chave ou até mesmo uma busca avançada que permite o filtro por seções, data, busca por títulos ou conteúdo, entre outros. As seções básicas do DOU consistem em:

1. Leis, decretos, resoluções, instruções normativas, portarias e outros atos normativos de interesse geral;
2. Atos de interesse dos servidores da Administração Pública Federal;
3. Extratos de instrumentos contratuais (acordos, ajustes, autorizações de compra, contratos, convênios, ordens de execução de serviço, termos aditivos e instrumentos congêneres) editais de citação, intimação, notificação e concursos públicos, comunicados, avisos de licitação entre outros atos da administração pública decorrentes de disposição legal.

Por conter tantas informações relevantes, o DOU é utilizado como fonte em diferentes níveis de pesquisa, principalmente para recuperação de informações. Guerra et al. (2015), por exemplo, analisa a resposta das agências reguladoras no combate a COVID-19 através da busca de conteúdo no DOU e também das 11 agências reguladoras no país. Outros estudos contemplam a análise regulatória de medicamentos no Brasil (Feitoza-Silva et al., 2017) e a inferência sobre risco de fraude nos contratos públicos (Lima et al., 2020)

No entanto, poucos estudos mencionam Diários Oficiais Municipais. Xavier et al. (2015) propuseram uma arquitetura híbrida de indexação do Diário Municipal de Cachoeiro de Itapemirim - ES.

No cenário da administração pública municipal, o setor do Diário Oficial Municipal tem a atribuição de indexar e publicar todo o conteúdo gerado pelo Poder Executivo através dos Atos Normativos contidos em

uma edição do Diário Oficial. Trata-se de uma tarefa nada trivial que consome tempo e recursos do setor, e quase sempre é auxiliada por ferramentas computacionais não específicas, o que compromete ainda mais a eficiência do processo. Extrair informações dos diários oficiais municipais pode ser muito útil para avaliação de políticas públicas, jornalismo, ou até mesmo como auditoria do governo.

Neste sentido, o projeto Querido Diário já mapeou quase todos os diários oficiais dos 5.570 municípios brasileiros, obtendo raspadores para mais de 2.300 municípios, cerca de 41% do total. A Open Knowledge Brasil junto com uma equipe de pessoas cientistas e engenheiras de dados e uma grande comunidade de colaboradores voluntários que trabalha no Querido diário alcançaram também com que qualquer formato de diário (PDF, Docx, RTF, HTML, JPEG, etc.) em .txt e fazer a extração e validação de CPFs e CNPJs dos diários oficiais. De modo a ampliar o projeto surge o Diário do Clima que foi um dos vencedores do edital da Google News Initiative (GNI). A GNI visa promover um jornalismo duradouro e diverso, no combate a desinformação em prol da transparência de dados.

O projeto Diário do Clima tem como objetivo gerar uma plataforma de consulta sobre iniciativas e políticas de combate à mudanças climáticas nos 5.570 municípios brasileiros no qual culminará em uma importante fonte para jornalistas. Para isso será necessário segmentar informações sobre políticas e iniciativas de combate à mudanças climáticas nos diários oficiais dos municípios brasileiros através de mecanismos de inteligência artificial.

Principais resultados atuais...

Explorando o Querido Diário

Por lei, todo ato público, seja do Poder Executivo, seja do Legislativo ou do Judiciário, deve ser publicado em um diário oficial. Contudo, não existe padronização de diários oficiais no Brasil nem uma central de dados que permita que a sociedade civil acesse o conteúdo de diários de maneira fácil e rápida. Para libertar os dados dos diários municipais, a Open Knowledge Brasil (OKBR) criou o projeto “Querido Diário”, cujo objetivo é mapear, coletar e processar todos os diários municipais do país e disponibilizá-los em formato aberto para a sociedade civil.

Para mapear diários oficiais municipais e extrair informações relevantes foi necessário compreender qual é a dimensão e escopo de um diário oficial municipal. Segundo o Diário Oficial do município de São Paulo o jornal é dividido em oito seções: Gabinete do Prefeito, Secretarias, Servidores, Concursos, Editais, Licitações, Câmara Municipal e Tribunal de Conta Um diário municipal está, teoricamente, categorizado logicamente em algumas seções. Essas seções correspondem a exigências legais de publicação de atos públicos. Contudo, como os municípios não seguem padrão uniforme de divulgação de informações, essas seções podem não ter títulos comuns ou mesma estrutura e nem são publicadas em todos os diários – afinal, nem sempre todos os atos públicos cuja publicação é exigida por lei ocorrem todos os dias. De modo geral, as seções obrigatórias são:

- Atos Normativos: Obrigações ou direitos aos cidadãos fruto de atividade do poder executivo (ou legislativo) municipal.
- Atos de Pessoal: Mudanças nas atividades dos funcionários municipais, contemplando ingresso em carreira pública, exoneração, concessão de afastamento temporário, etc.
- Contas Públicas: Obrigações de dívida dos municípios e peça orçamentária em respeito à LRF.
- Audiências Públicas: Informes para a população sobre audiências públicas referentes à atividade do governo municipal e implementação de políticas públicas locais.
- Licitações: Informes sobre os principais instrumentos de contratação de bens e serviços privados no Brasil pelo setor público.
- Processos Disciplinares: Informes de processos investigativos e disciplinares sobre conduta de servidores públicos ou contratantes do poder executivo municipal.
- Processos Tributários: Informes de acordos entre o fisco municipal e os contribuintes em processos administrativos tributários que afetam a arrecadação do município.
- Conselhos Municipais: Informes sobre a fiscalização de políticas públicas por parte da sociedade civil.

Como o interesse da sociedade civil por conteúdo de diários oficiais normalmente é focado em um tema, a segmentação do diário surge como nossa primeira prioridade de ciência de dados, uma vez que nos permitiria

“fatiar” o conteúdo e permitir buscas apenas em seções desejadas. Portanto, uma tarefa fundamental para a equipe da OKBR seria uma forma automatizada de segmentação do diário nestas categorias.

Parágrafo sobre os diferentes projetos: Histórico (o que veio com a pandemia/Fundação Lemann/Lemann - Ensino remoto/Diário do Clima) O projeto Querido Diário não surgiu do nada, ele veio inspirado pela Operação Serenata de Amor que criou a Rosie, uma robô que através de inteligência artificial consegue rastrear irregularidades nos gastos do Congresso Nacional. E também do Open Data Index A OKBR já tinha uma ideia de acessibilizar os diários oficiais e com tantas outras conquistas foi possível aproveitar o conhecimento acumulado para criar o projeto Querido Diário. Atualmente, a equipe da OKBR também atua em outros projetos e desde o início da pandemia da COVID-19 foram realizados projetos voltados para solucionar problemas relacionados a transparência de dados.. O projeto ‘Sorria #Estamos Vigando, Monitor da INDA, Transparência COVID-19.

Com uma equipe qualificada experiência e capacidade o Querido Diário compila os diários oficiais de todos os municípios brasileiros através de inteligência artificial. O projeto possui código aberto, ou seja, qualquer pessoa pode contribuir no projeto ou entender como a programação é arquitetada. De maneira simplificada os diários oficiais são coletados dos seus respectivos endereços eletrônicos e submetidos a raspadores (*spiders*) em que filtram o resultado para extrair informações como o formato do arquivo, tamanho e ser convertido em um documento acessível de PDF, concomitantemente são extraídas outras informações do arquivo como data, seção do diário que são armazenadas em uma base de metadados. ... Onde fica o arquivo bruto são extraídos o texto em formato TXT em formato aberto. Onde estão os metadados é realizado uma busca textual com o mecanismo de busca de análise de dados ElasticSearch, onde é possível indexar as informações dos diários e armazená-las centralmente facilitando futuras buscas. Após isso, uma interação entre os indexadores do ElasticSearch e o texto puro é realizada passando estes dados para a Interface de Programação de Aplicações (API) (usado pela máquina) que é integrada a uma plataforma de visualização para o usuário.

- Mapeamento dos diários
- Raspagem e estruturação dos dados
- Utilização e análise do produto

Diário do Clima

A partir do prêmio da Inovação promovido pela *Google News Initiative* foi possível o início do Projeto Diário do Clima que visa monitorar os atos de autorização e normativos relacionados a meio ambiente e clima publicados diariamente nos diários oficiais em todos os municípios brasileiros. A partir dos passos avançados pelo Querido Diário é possível aplicar algumas das técnicas previamente utilizadas na plataforma Diário do Clima. Para o projeto ser possível é necessário raspar e processar os documentos, com a utilização de mecanismos de inteligência artificial para categorizar cada ato e organizá-los por entidades relevantes permitindo a filtragem e priorização da informação. O Diário do Clima será uma plataforma que ajudará principalmente jornalistas e comunicadores.

O maior desafio da solução é selecionar os diários oficiais que tenham informações relevantes no âmbito do jornalismo ambiental e climático. Por isso essa iniciativa possui a colaboração de sete organizações: ((o))Eco, Eco Nordeste, Agência Envolverde, InfoAmazonia, ClimaInfo, #Colabora e Política por Inteiro. O objetivo final do projeto é ter uma plataforma aberta em que possa servir de consulta para rastrear o que foi desenvolvido nos municípios quanto à temática ambiental para ter controle e manter transparência nos atos.

A principal diferença dessa plataforma com a que foi criada pelo Querido Diário diz respeito ao escopo. O Querido Diário buscava filtrar os diários municipais sem conteúdo estruturado, em contrapartida o Diário do Clima busca concentrar em uma plataforma de visualização apenas diários oficiais que tenham conteúdos onde hajam informações relevantes relacionadas ao meio ambiente e clima nos municípios brasileiros. Dessa maneira será possível manter um banco de informações confiável e acessível corroborando com a proposta da Google News Initiative construindo um ecossistema de notícias mais sustentável, diversificado e inovador. Por trás do projeto trabalha uma imensa equipe formada desde cientista de dados a jornalistas, criando uma plataforma plural e acessível.

- Quais informações serão coletadas?

- Escopo/objetivo do diário do clima.

Marcos regulatórios - Meio Ambiente

No que diz respeito aos municípios na Constituição Federal de 1988: - Art. 24 Inciso 3º Inexistindo lei federal sobre normas gerais, os Estados exercerão a competência legislativa plena, para atender a suas peculiaridades - Art. 30. Compete aos Municípios: I - legislar sobre assuntos de interesse local; II – suplementar a legislação federal e a estadual no que couber (...) - Arts. 34 e 35: Das intervenções feitas pela União;

Um dos principais marcos ambientais dentro da legislação brasileira foi A Política Nacional do Meio Ambiente (Lei 6.938/81) que visa o desenvolvimento sócio-econômico atendendo aos princípios de ação governamental, racionalização, planejamento, proteção, controle e zoneamento das atividades, proteção e educação (Art. 2º). As diretrizes da Política Nacional do Meio Ambiente são formuladas em normas e planos, destinados a orientar a ação dos Governos da União, dos Estados, do Distrito Federal, dos Territórios e dos Municípios no que se relaciona com a preservação da qualidade ambiental e manutenção do equilíbrio ecológico, observados os princípios estabelecidos no art. 2º desta Lei (Art. 5º).

- SISNAMA
 - Órgão Superior: Assessora Presidente da República na formulação de políticas (CSMA - Conselho Superior do Meio Ambiente)
 - * Órgão consultivo e deliberativo: CONAMA - determina, quando julga necessário, a realização de estudos das alternativas e das possíveis consequências ambientais de projetos públicos ou privados, requisitando aos órgãos federais, estaduais e municipais, bem assim a entidades privadas, as informações indispensáveis para apreciação dos estudos de impacto ambiental, e respectivos relatórios, no caso de obras ou atividades de significativa degradação ambiental, especialmente nas áreas consideradas patrimônio nacional. **-PODER DE LEGISLAR**
 - * Órgão central: Secretaria do Meio Ambiente da Presidência da República - Ministério do Meio Ambiente;
 - * Órgãos executores: IBAMA e Instituto Chico Mendes
 - * Órgãos setoriais: órgãos ou entidades integrantes da Administração Pública Federal, direta ou indireta, bem como as fundações instituídas pelo Poder Público, cujas entidades estejam, total ou parcialmente, associadas às de preservação da qualidade ambiental ou de disciplinamento do uso de recursos ambientais
 - Órgãos Seccionais: órgãos ou entidades estaduais responsáveis pela execução de programas, projetos e pelo controle e fiscalização de atividades capazes de provocar a degradação ambiental. (Secretarias Estaduais de Meio Ambiente, IMA, IAP, CETESB, Inea, Polícia Militar Ambiental)
 - Órgãos Locais: órgãos ou entidades municipais, responsáveis pelo controle e fiscalização dessas atividades, nas suas respectivas jurisdições
- Fica alterado:

Art. 3º da Lei nº 10.165, de 27 de dezembro de 2000: A Lei nº 6.938, de 1981, passa a vigorar acrescida dos seguintes artigos: "Art. 17-Q. É o Ibama autorizado a celebrar convênios com os Estados, os Municípios e o Distrito Federal para desempenharem atividades de fiscalização ambiental, podendo repassar-lhes parcela da receita obtida com a TCFA." **Transferência de recurso com pessoal**

Instrumentos da PNMA:

- Padrões de qualidade ambiental;
- Zoneamento ambiental;
- Avaliação de impactos ambientais;
- Licenciamento e revisão de atividades efetiva ou potencialmente poluidoras;
- Incentivos à produção e instalação de equipamentos e a criação ou absorção de tecnologia, voltados para a melhoria da qualidade ambiental;

- criação de espaços territoriais especialmente protegidos pelo Poder Público federal;
- Sistema nacional de informações sobre o meio ambiente;
- Cadastro Técnico Federal de Atividades e Instrumentos de Defesa Ambiental;
- penalidades disciplinares ou compensatórias ao não cumprimento das medidas necessárias à preservação ou correção da degradação ambiental;
- instituição do Relatório de Qualidade do Meio Ambiente, a ser divulgado anualmente pelo IBAMA;
- garantia da prestação de informações relativas ao Meio Ambiente, obrigando-se o Poder Público a produzi-las, quando inexistentes.
- Cadastro Técnico Federal de atividades potencialmente poluidoras e/ou utilizadoras dos recursos ambientais.
- instrumentos econômicos, como concessão florestal, servidão ambiental, seguro ambiental e outros.

Com a promulgação da Constituição Federal em 1988, o meio ambiente passou ser direito difuso e coletivo.

Meio Ambiente na Constituição Federal:

- Art. 225;
- [Lei de Crimes Ambientais (1998)][http://www.planalto.gov.br/ccivil_03/LEIS/L9605.htm]: Art. 73. Os valores arrecadados em pagamento de multas por infração ambiental serão revertidos ao Fundo Nacional do Meio Ambiente, criado pela Lei nº 7.797, de 10 de julho de 1989, Fundo Naval, criado pelo Decreto nº 20.923, de 8 de janeiro de 1932, fundos estaduais ou municipais de meio ambiente, ou correlatos, conforme dispuser o órgão arrecadador.

obrigatoriedade de que todos os as propriedades e posses rurais do País façam parte do Sistema Nacional de Cadastro Ambiental Rural (Sicar)

Art. 3o da Lei nº 10.165, de 27 de dezembro de 2000: A Lei nº 6.938, de 1981 passa a vigorar acrescida dos seguintes Anexos VIII (atividades potencialmente poluidoras) e IX:(TCFA)

Órgãos

MMA: criado em 1992 (marco da Rio 92)

- Lei das águas (1997):
- Lei de Crimes Ambientais (1998):
- Política Nacional de Educação Ambiental (1999):
- Sistema Nacional de Unidades de Conservação (SNUC)(2000):
- Lei de Gestão de Florestas Públicas (2006):

Planos e Programas do MMA

- [Programa Floresta+][<https://www.gov.br/mma/pt-br/assuntos/servicosambientais/florestamais/ProgramaFloresta.pdf>];
- [Plano Nacional para Controle do Desmatamento Ilegal e Recuperação da Vegetação Nativa][<https://www.gov.br/mma/pt-br/assuntos/servicosambientais/controle-de-desmatamento-e-incendios-florestais/PlanoNacionalparaControledoDesmatamento20202023.pdf>] existem diretrizes para os municípios, consequentemente no [Plano Operativo 2020 - 2023][<https://www.gov.br/mma/pt-br/assuntos/servicosambientais/controle-de-desmatamento-e-incendios-florestais/PlanoOperativo20202023.pdf>] - Implementar programas e projetos de Pagamentos por Serviços Ambientais (PSA) existe responsabilidade das prefeituras municipais;

- Nenhum dos municípios descritos [aqui][<https://www.gov.br/mma/pt-br/assuntos/servicosambientais/controle-de-desmatamento-e-incendios-florestais/pdf/Listagemmunicipiosprioritriosparaesdepreveno2021.pdf>] estão na amostra inicial.
- [PLANAVEG Plano Nacional de Recuperação da Vegetação Nativa][https://www.gov.br/mma/pt-br/assuntos/servicosambientais/ecossistemas-1/planaveg_plano_nacional_recuperacao_vegetacao_nativa.pdf] - Entidades municipais como Secretarias de Meio Ambiente e Conselhos Municipais de Meio Ambiente para atuar na integração entre estâncias, sensibilização da população, e capacitação, bem como recuperação de áreas.
- [REDD++][http://redd.mma.gov.br/images/central-de-midia/pdf/publicacoes/notainformativa2018_captacaodescentralizacao.pdf];
- [ARPA, A3P, Cerrado Sustentável, Zoneamento ecológico, dentre outros][<https://antigo.mma.gov.br/programas-mma.html>];

IBAMA: Poder de polícia ambiental, responsável por recolher e repassar recursos da TCFA, ITR, CAR, APP, RL. objetivos institucionais relativos ao licenciamento ambiental, ao controle da qualidade ambiental, à autorização de uso dos recursos naturais e à fiscalização, monitoramento e controle ambiental e ações supletivas de competência da União, conforme legislação ambiental. Zoneamento e avaliação de impacto ambiental. CTF

- APP: Área de preservação permanente
- AUR: Área de Uso Restrito
- RL: Reserva Legal
- CAR: Cadastro Ambiental Rural é o registro público eletrônico de âmbito nacional, obrigatório para todos os imóveis rurais, com a finalidade de integrar as informações ambientais das propriedades e posses rurais referentes às APP, AUR, RL, de remanescentes de florestas e demais formas de vegetação nativa, e das áreas consolidadas, compondo base de dados para controle, monitoramento, planejamento ambiental e econômico e combate ao desmatamento.
 - * PRA: Programa de Apoio e Incentivo à Conservação do Meio Ambiente e aos Programas de Regularização Ambiental
 - * ITR: O CAR gera créditos tributários por meio de deduções na base de cálculo do Imposto Territorial Rural
- CTF: Cadastro Técnico Federal identifica as pessoas físicas e jurídicas sob controle ambiental e fiscalização ambiental, conforme previsto em legislação federal ou de âmbito nacional, gerando informações para a gestão ambiental no Brasil.
 - * TCFA: Taxa de Controle e Fiscalização Ambiental é um tributo para controle e fiscalização das atividades potencialmente poluidoras e utilizadoras de recursos naturais

ICMBio: apresentar e editar normas e padrões de gestão de Unidades de Conservação (UC) federais; propor a criação, regularização fundiária e gestão das UC federais; e apoiar a implementação do Sistema Nacional de Unidades de Conservação (SNUC). Contribui para a recuperação de áreas degradadas em UC. Fiscaliza e aplica penalidades administrativas ambientais ou compensatórias aos responsáveis pelo não cumprimento das medidas necessárias à preservação ou correção da degradação ambiental. Monitora o uso público e a exploração econômica dos recursos naturais nas UC onde isso for permitido, obedecidas as exigências legais e de sustentabilidade do meio ambiente. geração e disseminação sistemática de informações e conhecimentos relativos à gestão de UCs, da conservação da biodiversidade e do uso dos recursos faunísticos, pesqueiros e florestais. Contribui para a implementação do Sistema Nacional de Informações sobre o Meio Ambiente (Sinima) e aplica, no âmbito de sua competência, dispositivos e acordos internacionais relativos à gestão ambiental. Elabora o Relatório de Gestão das UC.

- UC: Unidade de Conservação

Setor privado

- Principal: ISO da série 14000;

- Outros: Certificação Orgânica; Comércio Justo; EurepGap (Boas Práticas Agrícolas (GAP); Certificação Socioambiental e Produção Integrada

Metodologia

A metodologia de trabalho foi sub-dividida em 5 etapas, sendo: Pré-processamento, mecanismos de busca, stopwords e stoplist, refinamento e avaliação de precisão.

1 Pré-processamento

Processo realizado previamente pelo Querido Diário, consiste em mapear os Diários Oficiais Municipais, obter raspadores para cada um destes e ser capaz de converter os documentos em texto puro e metadados, armazenando através do Elasticsearch.

2 Mecanismos de Busca

A recuperação de informações (RI) pode ser feita através de um sistema que funciona a partir de diferente mecanismos, sejam análise textual, filtragem de informação por meio da extração de stopwords, técnicas de redução de palavras a seus radicais (stemming), técnicas de indexação, arquivo invertido, modelos matemáticos e estatísticos para a representação de documentos, de consultas e a definição de coeficientes de similaridades, estruturas de categorização e de expansão de consultas por meio da utilização de Thesaurus (Aires, 2005). Para este projeto foram utilizados seis tipos de busca, a saber:

Intervalar

Retorna documentos com base na ordem e proximidade dos termos correspondentes. A consulta de intervalo.
Exemplo:

Querystring

Aqui é utilizada uma sintaxe para analisar e dividir o conjunto de caracteres de consulta fornecida com o termo.
Exemplo:

Booleana

Uma consulta que corresponde a documentos correspondentes a combinações booleanas de outras consultas.
 - must: O termo deve, obrigatoriamente, aparecer no documento.
 - filter: O termos deve aparecer nos documentos correspondentes.
 - should: O termo deve aparecer no documento.
 - must_not: O termo não deve, obrigatoriamente, aparecer no documento.

Boosting

Retorna documentos que correspondem a uma consulta positiva enquanto reduz a pontuação de relevância de documentos que não correspondem.

Fuzzy

Retorna documentos que contêm termos semelhantes ao termo de pesquisa medido por uma distância de edição.

- Mudando um personagem (box → fox)
- Removendo um caractere (black → lack)
- Inserindo um personagem (sic → sick)
- Transpondo dois caracteres adjacentes (act → cat)

Para encontrar termos semelhantes, a consulta do tipo fuzzy cria um conjunto de todas as variações ou expansões possíveis do termo de pesquisa dentro de uma distância de edição especificada. A consulta retorna correspondências exatas para cada expansão.

Expressões regulares

Retorna documentos que contêm termos que correspondem a uma expressão regular. Uma expressão regular é uma maneira de combinar padrões em dados usando caracteres de espaço reservado, chamados de operadores. Para obter uma lista de operadores suportados pela consulta regexp.

3 Stopwords e Stoplists

O processo de criação das palavras de busca é acompanhado de um longo trabalho envolvendo uma curadoria de jornalistas especializados. Foram selecionados voluntariamente 7 jornalistas para compor a equipe de curadoria e avaliar os documentos coletados através dos mecanismos de busca e avaliar qualitativamente os resultados. No final da 1ª etapa é previsto que cada um dos curadores avalie cerca de 768 termos de busca.

4 Refinamento

Após a coleta dos documentos e definição dos termos de busca será feita uma nova rodada de coleta de documentos para refinar os processos de busca. Essa etapa é necessária para deixar a plataforma mais eficaz e amigável, bem como melhorar o processamento dos documentos.

5 Avaliação de precisão

Análise de dados.. - Índices positivos e negativos; - Relação Qualidade do documento do ElasticSearch X Curadoria; - Qualidade dos highlights. Seguindo a equação 1

$$Precisão = \frac{n_{\text{de documentos relevantes}}}{n_{\text{de documentos recuperados}}}$$

Cronograma do projeto