

CONCURSO PRÊMIO FENAFIM/SENAM – 2026

CATEGORIA: PROFISSIONAL

Qualificação cadastral e crédito tributário municipal: OCR e reconhecimento de entidades aplicados a matrículas imobiliárias

RESUMO

Este trabalho apresenta o Busca Certidões, plataforma modular desenvolvida para qualificar o cadastro imobiliário e apoiar a constituição e a cobrança do crédito tributário municipal. A solução combina OCR, NER, validação de dados e mecanismo de busca textual, e é executada em infraestrutura própria e com ferramentas de código aberto. A pesquisa, de natureza aplicada, combinou procedimento experimental, para comparação de modelos de OCR, e estudo de caso em determinada municipalidade. Na avaliação de quatro modelos, o GLM-OCR apresentou maior precisão, enquanto o PaddleOCR-VL-1.6 obteve melhor eficiência computacional. No processamento de 8.004 certidões de matrículas, 6.723 foram vinculadas ao cadastro municipal, 5.201 geraram novos dados cadastrais e 1.347 instruíram a correção de divergências de sujeição passiva. Os resultados indicam que a solução amplia a qualidade das informações cadastrais e, ao incidir sobre a determinação da matéria tributável e a identificação do sujeito passivo, elementos privativos do lançamento nos termos do art. 142 do Código Tributário Nacional, contribui para a higidez do ato e para a segurança jurídica na cobrança do crédito tributário municipal.

Palavras-chave: Administração Tributária Municipal; Cadastro Imobiliário; Certidões de Matrículas de Imóveis; Reconhecimento Óptico de Caracteres; Reconhecimento de Entidades Nomeadas.

1. INTRODUÇÃO

Na tributação imobiliária municipal, a precisa determinação dos elementos cadastrais definidores do valor venal dos imóveis, a correta identificação do sujeito passivo e o historiamento das cadeias de transmissão dos bens e dos direitos a eles relativos são pressupostos da tributação do Imposto sobre a Propriedade Predial e

Territorial Urbana (IPTU) e do Imposto sobre Transmissão Inter Vivos de Bens Imóveis e de Direitos a eles Relativos (ITBI).

Estas informações estão presentes nos documentos cartorários e são dotadas de fé pública e de segurança jurídica. Sem lastro documental, o Município lança sobre base incerta e cobra de sujeito passivo que pode não ser o devido.

Neste sentido, a sistematização dos dados cartorários permite ao fisco validar as informações do cadastro municipal, obter as que nele faltam e direcionar ações fiscais voltadas à omissão de fatos geradores e à correção de lançamentos inconsistentes, com efeito direto sobre a cobrança administrativa e judicial. A literatura sobre a arrecadação do IPTU aponta a desatualização cadastral como um dos fatores que reduzem o potencial arrecadatário do imposto (CARVALHO JR., 2006).

Tal sistematização, porém, constitui um esforço não trivial, pois, embora muitos documentos sejam nativamente digitais, eles são frequentemente compartilhados entre diferentes órgãos em formato de imagem e com falhas. Uma parcela do acervo de certidões de matrículas cartorárias imobiliárias, por exemplo, é composta por imagens de documentos de qualidade visual tão baixa que até mesmo a leitura humana se torna um árduo exercício.

Agrava o quadro o fato de essas imagens serem apenas reproduções visuais dos respectivos originais: estáticas, não pesquisáveis e não estruturáveis, isto é, incapazes de ter o seu conteúdo organizado em campos e relações que sistemas possam reconhecer e processar.

A consequência é operacional, porque cada documento precisa ser buscado, conferido e interpretado manualmente, o que aumenta o tempo de resposta às

demandas dos contribuintes e restringe integrações, análises tributárias e ações fiscais em lote.

Superar essa limitação exige converter os documentos em textos pesquisáveis, reconhecer seu conteúdo e organizá-lo em campos específicos e reutilizáveis, de modo que dados relativos ao imóvel, aos proprietários e aos atos registrais possam ser pesquisados, cruzados e integrados aos sistemas tributários municipais, enquanto o documento original permanece preservado como fonte primária.

A pertinência dessa agenda extrapola o caso particular. A Lei Complementar nº 214/2025 tornou obrigatória a inscrição, pelas municipalidades, dos bens imóveis no Cadastro Imobiliário Brasileiro (CIB)¹, integrante do Sistema Nacional de Gestão de Informações Territoriais instituído pelo Decreto nº 11.208/2022 (BRASIL, 2022), cujos campos coincidem com os contidos nas certidões de matrículas imobiliárias. Municípios que não dispuserem de vínculo consistente entre a inscrição imobiliária e o registro cartorário transportarão para a base nacional as lacunas que hoje mantêm.

Nesse contexto, o objetivo geral deste trabalho é apresentar o desenvolvimento de uma plataforma modular, denominada Busca Certidões, para qualificar o cadastro imobiliário municipal e apoiar a constituição e a cobrança do crédito tributário. A solução é composta por um módulo de Reconhecimento Óptico de Caracteres (OCR), destinado ao processamento automatizado de certidões de matrículas imobiliárias, um módulo de Reconhecimento de Entidades Nomeadas (NER), voltado à extração de informações relevantes desses documentos, e uma

¹ O CIB integra o Sistema Nacional de Gestão de Informações Territoriais (SINTER), instituído pelo Decreto nº 11.208/2022, que reúne dados cadastrais, registrais, geoespaciais, fiscais e jurídicos relativos a bens imóveis. Os prazos de inscrição no CIB encerram-se em 31 de dezembro de 2025 para as capitais e o Distrito Federal e em 31 de dezembro de 2026 para os demais Municípios (BRASIL, 2022; 2025).

plataforma web construída essencialmente sobre bancos de dados não relacionais para a busca e recuperação de certidões com base em pesquisas textuais e nos dados reconhecidos durante o processamento.

Para alcançar esse objetivo, a pesquisa foi estruturada em objetivos específicos que abrangem quatro dimensões. No âmbito experimental, comparar quatro modelos de OCR de última geração, *open source* e gratuitos, segundo métricas de qualidade e desempenho, fundamentando a escolha da alternativa mais adequada às características dos documentos das matrículas imobiliárias. No âmbito do processamento de dados, disponibilizar o serviço de processamento documental e de reconhecimento de entidades, gerando conteúdo textual e dados cadastrais estruturados, por meio de uma API² executada em infraestrutura própria do órgão.

No âmbito funcional, oferecer mecanismo de busca que permita consultar tanto os dados extraídos quanto as próprias certidões. Por fim, no âmbito negocial, aferir o aproveitamento dos dados estruturados pelo cadastro imobiliário municipal e seus reflexos e potencialidades sobre a correta identificação do sujeito passivo e a cobrança do crédito tributário.

2. FUNDAMENTAÇÃO TEÓRICA

2.1. Qualificação cadastral e constituição do crédito tributário

O CTN determina:

Art. 142. Compete privativamente à autoridade administrativa constituir o crédito tributário pelo lançamento, assim entendido o procedimento administrativo tendente a verificar a ocorrência do fato gerador da obrigação correspondente, determinar a matéria tributável, calcular o montante do tributo devido, identificar o sujeito passivo e, sendo caso, propor a aplicação da penalidade cabível.

² Sigla que em português significa Interface de Programação de Aplicações. É uma maneira muito popular de se integrar sistemas computacionais.

O lançamento, no caso do IPTU, de acordo com SCHOUERI (2019), é *ex officio*, ou seja, a Administração Tributária encarrega-se de apurar o valor monetário e notificar o contribuinte a respeito. E este lançamento usa como referência o cadastro imobiliário. Daí emerge a grande importância de se ter um cadastro imobiliário higienizado e atualizado: são estas qualidades que garantem o lançamento correto, colaboram com a promoção da justiça fiscal, facilitam a cobrança do crédito e o atendimento ao contribuinte.

É por essa razão que as deficiências cadastrais tangenciadas neste trabalho não constituem mera imperfeição administrativa, uma vez que atingem a determinação da matéria tributável e a identificação do sujeito passivo, elementos que o art. 142 do CTN comete privativamente à autoridade lançadora.

2.2. Inteligência artificial e suas técnicas no contexto cadastral

A inteligência artificial (IA) é uma disciplina que mescla conhecimentos de diferentes áreas, em especial da ciência da computação e estatística, que se propõe a simular funções cognitivas humanas, como o aprendizado e a capacidade de reconhecer padrões.

Buarque (2023) expõe que a IA é o centro das atenções, mas alerta que apenas uma parte menor dos problemas computacionais é adequada a ela. Entre diferentes classes de problemas aptas à IA apresentadas pelo ilustre professor, temos a dos que não têm solução conhecida ou que é custosa. E neste grupo se encontram os desafios impostos para se atingir o objetivo de negócio deste trabalho: Reconhecimento Óptico de Caracteres (OCR) e o Reconhecimento de Entidades Nomeadas (NER, em inglês), pois elas, em conjunto, habilitam a extração de dados dos documentos cartorários em formato de texto pesquisável e estruturado e também a inserção no cadastro imobiliário em escala e a baixo custo.

2.2.1. Reconhecimento Óptico de Caracteres (OCR)

O OCR é um conjunto de tecnologias que converte informações visuais de documentos digitalizados ou capturados por dispositivos ópticos em textos editáveis e processáveis por sistemas computacionais (KASTURI; O’GORMAN; GOVINDARAJU, 2002). Ao converter imagens de documentos em texto legível por máquina, o OCR torna informações antes restritas a documentos físicos ou arquivos de imagem disponíveis para aplicação de técnicas de recuperação de informação, processamento de linguagem natural e análise automatizada (NGUYEN et al., 2021).

2.2.2. Reconhecimento de Entidades Nomeadas (NER)

NER é um ramo do Processamento de Linguagem Natural (PLN) voltado à identificação e à classificação de menções a entidades em textos não estruturados. O objetivo consiste em localizar trechos que correspondem a categorias previamente definidas, como pessoas, organizações e localizações, além de expressões temporais e numéricas, como datas, horários, valores monetários e percentuais, atribuindo-lhes rótulos semânticos conforme o esquema de anotação adotado (SONKAR; WANG; BARANIUK, 2023). Assim, o NER permite que as entidades identificadas sejam classificadas e seus dados sejam estruturados, possibilitando a higienização e o enriquecimento de bases de dados, a integração com outros sistemas e a automação de fluxos de trabalho.

3. METODOLOGIA

Quanto à natureza, o presente trabalho caracteriza-se como uma pesquisa aplicada, pois busca produzir conhecimentos utilizáveis na resolução de um problema específico (PRODANOV; FREITAS, 2013, p. 51). Quanto aos procedimentos, combina pesquisa experimental, mediante testes controlados de

desempenho dos modelos, e estudo de caso, ao analisar documentos e fluxos reais da administração pública municipal (PRODANOV; FREITAS, 2013, p. 57, 60).

Para alcançar os objetivos propostos, o trabalho foi desenvolvido em nove etapas: identificação e delimitação do problema; seleção dos modelos de OCR; composição da amostra documental; elaboração das transcrições de referência; execução dos testes; cálculo das métricas; escolha do modelo mais adequado; desenvolvimento da API; e desenvolvimento de uma plataforma web para busca de certidões. A parte tecnológica foi elaborada com a premissa de usar softwares gratuitos e *open-source*, sem custo de investimento e nem de operação extras e capazes de serem executados em infraestrutura de TI própria e sem demandar computadores sofisticados.

Para avaliar os modelos, foram produzidas transcrições manuais de documentos reais, utilizadas como referência para a comparação com os resultados gerados. Para representar a variabilidade documental, foram selecionadas quatro certidões de matrículas, produzidas em diferentes décadas, totalizando 25 páginas e cerca de 7.334 palavras, com variações de conservação, nitidez, densidade textual e elementos adicionais, como assinaturas e trechos riscados.

Foram avaliados quatro modelos de IA para OCR e análise documental: GLM-OCR, PaddleOCR-VL-1.6, Unlimited-OCR e DeepSeek-OCR-2. Os experimentos foram realizados em um computador de uso pessoal, com processador Apple M5, com 16 GB de memória unificada e GPU integrada. A avaliação de precisão das transcrições utilizou as métricas Taxa de Erro de Caracteres (CER) e Taxa de Erro de Palavras (WER). Já para avaliar a eficiência computacional, foram utilizadas as métricas de Tempo Médio de Processamento por Página (TMPP) e Consumo de Memória (CM). A seleção priorizou baixas taxas de

erro, estabilidade em documentos degradados e baixo consumo de memória, essencial ao ambiente de TI limitado do órgão.

Após a extração textual, foi desenvolvido um módulo de Reconhecimento de Entidades Nomeadas (NER) com o modelo Gemma 4, de 4 bilhões de parâmetros. O *pipeline* de inferência submeteu os textos ao modelo para identificar, classificar entidades, como número de documentos de pessoas e de imóveis e área dos imóveis, convertendo o conteúdo não estruturado em dados organizados para uso nas funcionalidades de busca e na integração com sistemas internos.

Por fim, as APIs de OCR e NER foram implementadas com Django REST Framework (DRF), e a plataforma web foi implementada em Django, utilizando o Elasticsearch³ como mecanismo de busca. A arquitetura e o funcionamento desses componentes são detalhados no capítulo seguinte.

4. RESULTADOS

4.1. Avaliação de desempenho dos modelos de OCR

A Tabela 1 apresenta o comparativo de desempenho dos quatro modelos avaliados. Todos foram submetidos às mesmas condições de teste, o que permite uma comparação consistente. Para avaliar a precisão, foram consideradas as métricas taxa de erro de caracteres (CER) e taxa de erro de palavras (WER). Para eficiência computacional, foram utilizadas as métricas de tempo médio de processamento por página (TMPP) e consumo de memória (CM).

Tabela 1: Comparativo de Desempenho dos Modelos de OCR

Modelo	CER	WER	TMPP	CM
DeepSeek-OCR-2	69,1%	141,5%	29,1 s	10 GB
GLM-OCR	6,4%	14,9%	20,1 s	5,02 GB
PaddleOCR-VL-1.6	9,8%	18,8%	8,75 s	4,13 GB

³ É um motor de busca de dados baseado em documentos e muito rápido.

Unlimited-OCR	59,2%	109,2%	20,6 s	10 GB
---------------	-------	--------	--------	-------

Fonte: Elaborado pelos autores.

Em precisão, o GLM-OCR apresentou o melhor desempenho, com os menores valores de CER (6,40%) e WER (14,90%), seguido pelo PaddleOCR-VL-1.6 (9,80% e 18,80%, respectivamente). Embora tenham aproximadamente o triplo do porte desses modelos, o DeepSeek-OCR-2 e o Unlimited-OCR registraram taxas de erro muito mais altas: 69,10% e 141,50% no primeiro caso, e 59,20% e 109,20% no segundo.

Em eficiência computacional, o PaddleOCR-VL-1.6 foi o mais rápido (TMPP de 8,75 segundos) e consumiu menos memória (4,13 GB), enquanto o GLM-OCR teve TMPP de 20,1 segundos e pico de 5,02 GB. O DeepSeek-OCR-2 apresentou o maior tempo médio por página (29,1 segundos), e ele e o Unlimited-OCR atingiram 10 GB de pico de memória. Assim, o PaddleOCR-VL-1.6 oferece o melhor equilíbrio entre precisão e custo computacional, sendo mais indicado quando a prioridade é processar grandes lotes com menor uso de recursos.

Figura 1: Resultado de documento processado pelo GLM-OCR

DOCUMENTO ORIGINAL

NÚMERO DO TÍPOLO ANTERIOR: -Livro nº 02, de Registro Geral, Matrícula nº 3957, Ficha nº 3957, em data de 15.06.1983, e o desmembramento no AV-3 da mesma matrícula, neste 4º Registro Imobiliário da Capital.

TRANSCRIÇÃO COM GLM-OCR

NÚMERO DO TÍPOLO ANTERIOR: -Livro nº 02, de Registro Geral, Matrícula nº 3957, Ficha nº 3957, em data de 15.06.1983, e o desmembramento no AV-3 da mesma matrícula, neste 4º Registro Imobiliário da Capital.

Fonte: Elaborado pelos autores.

No entanto, o GLM-OCR apresentou as menores taxas de erro e mostrou-se estável mesmo em documentos com alto grau de degradação, conforme evidenciado pela Figura 1. Esse desempenho, associado ao baixo pico de memória, foi

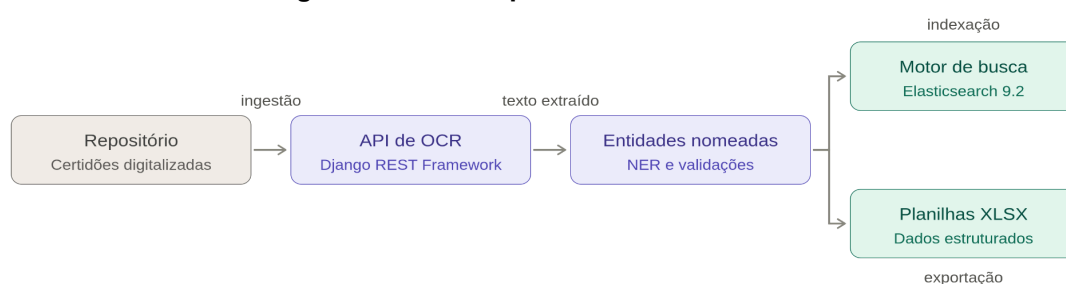
determinante para sua seleção como a alternativa mais adequada à aplicação proposta.

4.2. Visão tecnológica da solução

A solução foi estruturada em dois eixos independentes: i) o de recebimento de documentos, processamento e organização das informações; e ii) de sua disponibilização para consulta. Por utilizar tecnologias *open source* gratuitas e poder ser executada em computadores de baixo custo, a solução elimina a dependência de licenças proprietárias e amplia a viabilidade de implantação em outras Administrações Tributárias, mesmo aquelas com limitações orçamentárias.

O primeiro eixo, ilustrado na Figura 2, é responsável pelo fluxo de processamento dos documentos e armazenamento das informações extraídas. Na primeira etapa, os documentos cartorários (i.e., certidões de matrícula de imóveis) em formato de imagem ou PDF não pesquisável são processados através da API do OCR, baseado em IA, para transformar imagens em textos pesquisáveis e estruturáveis.

Figura 2: Fluxo de processamento de documentos



Fonte: Elaborado pelos autores.

O passo seguinte é a mencionada estruturação, efetuada a partir do emprego de técnicas de reconhecimento de entidades nomeadas e complementada com abordagens baseadas em expressões regulares.

Em seguida os dados passam por processos de validação por meio do cálculo de dígito verificador para minimizar a probabilidade de que a estruturação

tenha ocorrido de maneira imprecisa. Isso traz robustez ao resultado final. Após a validação, são geradas três planilhas em formato XLSX para facilitar a análise pelos auditores. A primeira, apresentada na Figura 3, com mascaramento das informações, apresenta dados de identificação pessoal e do imóvel.

Figura 3: Amostra de planilha com dados cadastrais de pessoas e imóveis estruturados

	A	B	C	D	E	F	G
1	Cartório	Matrícula	Chave	Valor	Válido?	Página	Ato registral
2	i_1	'1***1	SEQUENCIAL	'7*****2	SIM	3	7
3	i_1	'1***1	SEQUENCIAL	'7*****0	SIM	2	4
4	i_1	'7***3	CPF	'0*****5	SIM	1	1
5	i_1	'7***3	CPF	'5*****2	SIM	5	3
6	i_1	'7***6	CPF	'9*****0	NÃO	1	1
7	i_1	'1***1	CNPJ	'6*****4	SIM	4	10
8	i_1	'1***2	CNPJ	'0*****8	SIM	1	1
9	i_1	'8***1	DSOFL	'6*****5	NA	2	4
10	i_1	'1***2	INSCRICAO IMOB.	'6*****9	SIM	3	4

Fonte: Elaborado pelos autores.

A partir desta planilha é possível verificar diversos documentos encontrados e seus respectivos números por cartório e número de matrícula. A coluna “Válido?” indica se o valor foi aprovado na validação ou não. NA é não aplicável por não existir validação por dígito verificador para a entidade. Também é possível saber o número da página e o número do ato registral onde o dado foi encontrado⁴.

Figura 4: Amostra de planilha com dados de endereço semiestruturados

	A	B	C	D	E	F	G
1	Cartório	Matrícula	Chave	Valor	Válido?	Página	Ato registral
2				L:- Apartamento n° 102 (cento e dois), localizado no i1° pavimento elevado, do EDIFICIO A*A E****H, situado na Rua Dr. G**** R***** C****, n° 129, no bairro de Boa Viagem, freguesia dos Afogados, nesta cidade, com			
	i_1	1***6	Endereço		NA	1	1
3				L:- apartamento n° 301 (trezentos e um), localizado no 3° pavimento tipo, do Bloco 01, do EDIFICIO E***** J**** M***** F****, situado na Avenida B** V****, n° 998, no bairro de Boa Viagem, nesta cidade, com			
	i_1	1***7	Endereço		NA	1	1

Fonte: Elaborado pelos autores.

Conforme a Figura 4, a segunda planilha apresenta os dados de endereços dos imóveis semiestruturados. Já a terceira planilha, apresentada na Figura 5, permite a verificação da área e o seu tipo. O tipo de área é indicado pelo valor “True” das colunas *booleanas* (i.e., composta apenas de valores “True” ou “False”).

⁴ O conceito de ato registral empregado neste trabalho não é juridicamente preciso, mas foi utilizado por facilitar a organização das informações. Ele engloba anotações, averbações e retificações, sendo seu número dado pela ordem em que aparece na certidão.

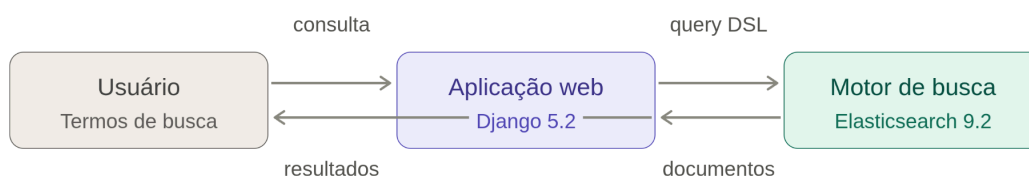
Figura 5: Amostra de planilha com dados de áreas estruturadas

Mat	Valor	Texto	Comum	Privativa	Construção	Condomínio	Própria	Real	Total
15	1****1	63,05 dependências de empregada, e com 4rea prépria de 63,05m	False	False	False	False	True	False	False
16	1****1	6,53 . area de condomínio de 6,53m?	False	False	False	True	False	False	False
17	1****2	55,00 as seguintes Areas: area toal	False	True	False	False	False	False	False
18	101295	51,75 Grea privativa, 55,00m estacionamento ti e tera as seguintes áreas: privativa de 51,75m	False	True	False	False	False	False	False

Fonte: Elaborado pelos autores.

Concluído o primeiro eixo, o texto pesquisável e os dados estruturados são indexados no Elasticsearch. A partir dessa indexação, o segundo eixo da solução realiza as consultas e disponibiliza os resultados ao usuário final, permitindo a busca e a recuperação das certidões, conforme ilustrado na Figura 6.

Figura 6: Fluxo para consulta e recuperação de certidões de matrículas



Fonte: Elaborado pelos autores.

A Figura 7 apresenta a interface do sistema, onde o usuário tem acesso à tela que possibilita inserir os termos a serem pesquisados, similarmente a buscas em ferramentas como o Google e o Yahoo.

Figura 7: Interface da plataforma web de busca de certidões

Busca Certidões

☒ 1º RGI
 ☒ 2º RGI
 ☒ 3º RGI
 ☒ 4º RGI
 ☒ 5º RGI
 ☒ 6º RGI
 ☒ 7º RGI

Quantidade de resultados: 33

Matrícula: 37195
 Cartório: 1. Seção: NI. Página: NI

Emissão: 2024-10-30

...comprimento de ambos os lados, confrontando-se pela frente com a rua Bartolomeu Bueno, antiga r: Um...

...pelo lado direito com o lote de terreno n: @Q, aves Ve, da rua Bartolomeu Bueno, pertencente a Rodoviar...

...mingos Ltda. ; pelo lado esquerdo com o lote de terren aay adrs IV, da rua Bartolomeu Bueno, pertencente...

...Bueno, na Imbiribeira, freguesia dos Alogados, nesta cidade, medindo 15,00m de frente por 35,00m de...

...IDENTIFICACAO DO IMOVEL:— Lote de terreno próprio nQ 10, de quadra IV, situado a rua Us, atual rua Bartolomeu...

Matrícula: 49398
 Cartório: 1. Seção: NI. Página: NI

Emissão: 2024-10-30

...Bueno, antiga ao arquivada neste ofício, em data de 24/03/88, d e tudo u fe...

...Projetada T, conf eemgs do art. 213, 8 19 da a correção parcial da nfrontações dos funda, do ua Bartolomeu...

Matrícula: 1594
 Cartório: 1. Seção: NI. Página: NI

Emissão: 2024-10-30

...Bueno, lado direito com o lote nQ 15, da m yadra, com Frente para a rua Bartolomeu Bueno, pertencente...

...situado em rua Projetada, paralela a Estrada da Imbiribeira, confrontando-se pela frente a rua Bartolomeu...

Fonte: Elaborado pelos autores.

A plataforma permite buscas do tipo *full text search*, mecanismo sofisticado capaz de apresentar resultados idênticos (i.e., exatos), mas também buscas que

combinam em algum grau com a intenção da busca (i.e., por aproximação). Além disso, o sistema permite filtrar os cartórios que serão pesquisados.

O tempo de resposta de uma pesquisa usualmente é menor do que um segundo para varrer mais de 60.000 certidões. Como resposta à busca, é apresentado um quantitativo de resultados encontrados e cada retorno da pesquisa apresenta: i) os trechos da certidão que correspondem aos termos pesquisados, ii) possibilidade de baixar a certidão matriz em formato pdf ou sua versão textual; e iii) quando a certidão foi obtida para facilitar eventuais atualizações.

4.3. Efeitos do processamento automatizado sobre os elementos do lançamento tributário

4.3.1. Cadastro fiscal e registro imobiliário: mecanismo para a vinculação entre bases

A desarticulação entre o cadastro fiscal e o registro imobiliário é problema estrutural reconhecido na literatura brasileira sobre cadastro territorial (CARNEIRO, 2003), e as diretrizes federais para o Cadastro Territorial Multifinalitário já apontavam a necessidade de integração entre essas bases.

A Administração Tributária e o registro imobiliário identificam o mesmo bem por chaves distintas. Na Administração Tributária Municipal, a chave de pesquisa é a inscrição imobiliária; no ofício de registro, é o número da matrícula. Enquanto não houver correspondência entre elas, cada consulta registral obriga um servidor a procurar o imóvel pelo endereço descritivo, um a um.

O cadastro do Município examinado reunia, em 2024, mais de 400 mil inscrições imobiliárias ativas, distribuídas por diversos ofícios de registro de imóveis. Desse contingente, 68,8% não possuíam nenhum dado proveniente de cartório de imóveis registrado. Para esse conjunto de imóveis, o Município lançou naquele

exercício R\$ 662,3 milhões de imposto sobre a propriedade e taxa de resíduos sólidos, contra R\$ 499,7 milhões dos imóveis com dados cartorários cadastrados.

Ou seja, 57% do lançamento anual recaía sobre imóveis cujo lastro registral o próprio fisco desconhecia. A diferença de comportamento entre os dois grupos acompanha essa assimetria: em 2024, os imóveis com dados cartorários cadastrados apresentavam adimplência de 88,7% no IPTU, contra 70,5% dos demais⁵.

O ganho estruturante da solução está na correspondência entre as duas bases. Estabelecida a correspondência entre base cartorária e cadastro imobiliário do município, o acervo registral passa a ser consultável pela chave já utilizada pelo fisco, ou seja, pelo número da inscrição municipal do imóvel. O módulo de reconhecimento de entidades chega a ela por dois caminhos: pelo endereço descritivo constante da matrícula ou, quando a inscrição municipal já foi averbada no documento, pela leitura direta dela.

4.3.2. Aferição do aproveitamento cadastral em lote

Foram processadas pela solução 8.004 certidões de matrículas provenientes de um dos escritórios de registro com atuação no município. Em seguida, o resultado foi validado matrícula a matrícula, em múltiplas etapas, por oito servidores em regime de dedicação parcial. A Tabela 2 apresenta os quantitativos de cada etapa.

Tabela 2: Funil de aproveitamento cadastral das matrículas processadas

Etapa	Atividade	Quantidade	% da etapa anterior	% do lote
1	Conferência do resultado do processamento das certidões de matrículas	8.004	—	100,00%
2	Identificação no cadastro dos imóveis presentes nas certidões de matrículas processadas	6.723	83,99%	83,99%

⁵ Dados extraídos de relatório de avaliação de projeto de qualificação cadastral, elaborado no âmbito de plano de equilíbrio fiscal firmado junto a organismo internacional de fomento (2021-2024).

3	Atribuição de dados cadastrais a imóveis que antes não os possuíam	5.201	77,36%	64,98%
4	Análise detalhada da sujeição passiva de imóveis	4.780	91,91%	59,72%
5	Identificação de imóveis com informação apta a corrigir a propriedade	1.347	28,18%	16,83%

Fonte: Elaborado pelos autores.

Documentos com marcas d'água, inclinação e degradação renderam associação confiável ao cadastro em pouco mais de oito de cada dez casos (etapa 2), e o resíduo de 1.281 certidões de matrículas não identificadas constitui contingente delimitado, apto a tratamento manual dirigido em vez de varredura integral.

Mais expressivo é o resultado da etapa 3: dos imóveis identificados, três quartos não possuíam qualquer informação registral no cadastro. Nesses casos, a operação criou informação onde havia lacuna, em vez de refinar dado preexistente. E, entre os casos submetidos a exame detalhado de titularidade (etapa 4), pouco mais de um em cada seis apresentou divergência ou lacuna sanável.

Relevante dimensionar o esforço operacional empregado na validação manual desse contingente de certidões de matrículas. Disponível o documento, identificar o imóvel consumiu cerca de três minutos de trabalho especializado, e a determinação do sujeito passivo, outros três. Aplicados aos volumes da Tabela 2, esses tempos somam aproximadamente 639 horas efetivamente despendidas: a validação conferiu a integralidade do lote, para garantir a acurácia dos resultados, eliminando qualquer viés que pudesse ser introduzido por técnicas de amostragem.

É importante ressaltar que toda a validação, inclusive a abordagem em múltiplas etapas, reproduziu exatamente o procedimento manual tradicionalmente utilizado, o que permite projetar com segurança o esforço equivalente em maior escala. Aplicado ao passivo cadastral atual do Município, esse percurso demandaria

décadas para a equalização integral, prazo que o processamento automatizado reduz à ordem de meses.

4.3.3. Sujeição passiva e cobrança do crédito tributário

A verificação de titularidade incide diretamente sobre a identificação do sujeito passivo, elemento que o art. 142 do CTN atribui privativamente à autoridade lançadora. O processamento produziu, nesse plano, três efeitos de impacto crescente sobre a qualidade do crédito e sobre a cobrança.

O primeiro diz respeito à correção da identificação do sujeito passivo, seja por desatualização, seja por incorreção propriamente dita do dado no cadastro municipal. O segundo é a obtenção de CPF ou CNPJ de titulares que constavam do cadastro apenas pelo nome. Esse ponto deixou de ser questão de qualidade de dado e virou condição de admissibilidade processual: a Resolução CNJ nº 617/2025, determina a extinção das execuções fiscais que não indiquem o CPF ou o CNPJ da parte executada. O alcance é nacional. Mais de três milhões de execuções fiscais sem essa informação estão sujeitas a extinção nas Justiças Estadual e Federal.

O terceiro efeito é a identificação de espólios, inventariantes e herdeiros em imóveis registrados em nome de titular falecido. Como o Código Tributário Nacional atribui ao espólio a responsabilidade pelos tributos devidos pelo *de cujus* até a abertura da sucessão (art. 131, III), identificar o representante legal é condição para a regular formação do polo passivo. No fluxo corriqueiro, porém, o óbito só aparece depois do ajuizamento, quando corrigir já significa emendar a inicial ou perder o feito; e cada crédito ajuizado contra sujeito passivo mal qualificado consome integralmente o custo do processo sem perspectiva de retorno⁶.

⁶ O custo mínimo estimado de uma execução fiscal, calculado sobre o valor da mão de obra, é de R\$9.277,00, e mais da metade delas é ajuizada por valor inferior a R\$10.000,00 (CONSELHO NACIONAL DE JUSTIÇA, 2024).

Além da amarração entre as chaves e da identificação do sujeito passivo, o processamento devolve ao cadastro parâmetros que, em parte considerável do acervo, carecem de validação ou atualização: lote, quadra e loteamento; área do lote, testada principal e área construída. Esses três últimos repercutem diretamente na avaliação do imóvel e, por consequência, na base de cálculo dos impostos sobre a propriedade e sobre a transmissão. Trata-se, em termos do art. 142, de deficiência na determinação da matéria tributável, vício que se manifesta como lançamento regular sobre base inferior à devida, e não como inadimplemento.

No âmbito da cobrança judicial, o vínculo entre inscrição imobiliária e matrícula instrui a execução fiscal e a eventual penhora do imóvel junto ao registro, conferindo à Procuradoria elementos de que frequentemente carece ao ajuizar. O contribuinte também se beneficia, uma vez que a certidão de matrícula é documento pago⁷, exigido na instrução de metade dos tipos de processos administrativos imobiliários que o cidadão pode abrir junto ao Município; quando o dado cartorário já integra o cadastro, o próprio Município a emite sem custo ao requerente.

5. CONSIDERAÇÕES FINAIS

A pesquisa teve por objetivo qualificar o cadastro imobiliário municipal e reforçar a hígidez do crédito tributário por meio da plataforma modular “Busca Certidões”, desenvolvida especificamente para automatizar o processamento, a extração e a recuperação de informações em certidões de matrículas imobiliárias. A solução envolveu a avaliação de modelos de OCR, APIs de OCR e NER, um mecanismo de busca baseado em banco de dados não relacional e a análise de seus efeitos sobre a gestão tributária municipal. Os resultados indicam que a

⁷ O custo por certidão de matrícula imobiliária varia entre R\$ 65 e R\$ 150 no mercado local em 2026.

arquitetura permite converter documentos analógicos em dados estruturados e pesquisáveis, apoiando consultas e atividades de fiscalização tributária.

No que tange aos resultados experimentais, os testes realizados confirmaram a eficácia da arquitetura proposta, com o GLM-OCR destacando-se como o modelo de melhor desempenho nas métricas de qualidade avaliadas. A API apresentou comportamento robusto durante o processamento documental, enquanto a plataforma de busca demonstrou desempenho consistente na recuperação das certidões.

Os resultados descritos são de um primeiro estágio. Não permitem afirmar que o projeto esteja concluído, mas que o percurso é viável, que seus resultados são mensuráveis e que ele ataca o limite do tratamento manual, isto é, o processamento em lote do acervo de certidões detido pelo órgão.

Cabe comparar esse caminho com a alternativa mais óbvia, que consiste em obter dos cartórios a base de dados estruturada, por convênio. É solução legítima e desejável, mas esbarra em firmar instrumento que respalde o compartilhamento com cada ofício jurisdicionante, estipular periodicidade e efetivamente receber os dados. Superadas essas etapas, a base de dados ainda tende a chegar como carga periódica, implicando em algum grau de defasagem. Dominar o tratamento da matrícula acrescenta o que a carga não oferece: processar os documentos tão logo cheguem à repartição, qualquer que seja o processo que os trouxe. As duas vias, portanto, se somam; não se substituem.

A capacidade de processar o documento no fluxo corrente, e não apenas em lotes retrospectivos, abre um campo de aplicação que este trabalho não esgota e que compreende a triagem automática de exigências documentais na abertura do processo, o enquadramento prévio do pedido do contribuinte e a revisão cadastral

de ofício disparada por documento já exigido. As três dependem da mesma capacidade aqui validada e guardam entre si ordem de precedência técnica, constituindo agenda de desenvolvimento cujo horizonte depende de decisão institucional de cada administração.

A inscrição dos imóveis no CIB, nos prazos fixados⁸ pela Lei Complementar nº 214/2025, exige mais do que cumprimento de calendário. Como a integração ao SINTER tende a reproduzir em escala nacional a qualidade das bases de origem, o Município que a ela chega com cadastro deficiente compromete, além do próprio lançamento, a confiabilidade do sistema territorial que a União se propõe a constituir.

Este trabalho demonstra que boa parte da informação registral necessária a essa integração já se encontra arquivada nos fiscos municipais, estagnada em documentos de imagem, e que seu tratamento por infraestrutura própria supre lacunas cadastrais com meios de que o órgão já dispõe.

Qualquer Administração Tributária Municipal com acervo digitalizado de certidões de matrículas cartorárias pode refazer o percurso. O que se oferece aqui é um método, com resultado medido, e por isso mesmo potencial de adaptação às realidades individuais de cada Fisco.

6. REFERÊNCIAS BIBLIOGRÁFICAS

BRASIL. **Decreto nº 11.208, de 26 de setembro de 2022.** Dispõe sobre o Sistema Nacional de Gestão de Informações Territoriais e regula o compartilhamento de dados relativos a bens imóveis. Brasília, DF: Presidência da República, 2022.

Disponível em:

https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2022/decreto/d11208.htm.

Acesso em: 1 ago. 2026.

⁸ O CIB integra o Sistema Nacional de Gestão de Informações Territoriais (SINTER), instituído pelo Decreto nº 11.208/2022, que reúne dados cadastrais, registrais, geoespaciais, fiscais e jurídicos relativos a bens imóveis. Os prazos de inscrição no CIB encerram-se em 31 de dezembro de 2025 para as capitais e o Distrito Federal e em 31 de dezembro de 2026 para os demais Municípios (BRASIL, 2022; 2025).

BRASIL. **Lei Complementar nº 214, de 16 de janeiro de 2025**. Institui o Imposto sobre Bens e Serviços, a Contribuição Social sobre Bens e Serviços e o Imposto Seletivo. Brasília, DF: Presidência da República, 2025. Acesso em: 1 ago. 2026.

BRASIL. **Lei nº 5.172, de 25 de outubro de 1966**. Dispõe sobre o Sistema Tributário Nacional e institui normas gerais de direito tributário aplicáveis à União, Estados e Municípios. Brasília, DF: Presidência da República, 1966. Acesso em: 1 ago. 2026.

BUARQUE, Fernando. **Aula 2**. [ANÔNIMO], 2023. 1 apresentação de slides. Notas de aula do módulo Inteligência Artificial do Programa de Formação de RH-Público em Inteligência Artificial, Ciência de Dados e Analytics (RH.IA-PE), 2023.2.

CARNEIRO, Andrea Flávia Tenório. **Cadastro imobiliário e registro de imóveis**. Porto Alegre: Sergio Antonio Fabris Editor, 2003.

CARVALHO JR., Pedro Humberto Bruno de. **IPTU no Brasil: progressividade, arrecadação e aspectos extrafiscais**. Texto para Discussão nº 1251. Brasília: IPEA, 2006.

CONSELHO NACIONAL DE JUSTIÇA. **Resolução nº 617, de 12 de março de 2025**. Altera a Resolução nº 547, de 22 de fevereiro de 2024. Brasília, DF: CNJ, 2025. Acesso em: 1 ago. 2026.

KASTURI, Rangachar; O'GORMAN, Lawrence; GOVINDARAJU, Venu. **Document image analysis: a primer**. Sadhana, v. 27, p. 3–22, 2002. DOI: 10.1007/BF02703309. Disponível em: <https://doi.org/10.1007/BF02703309>. Acesso em: 29 jul. 2026.

NGUYEN, Thi Tuyet Hai; JATOWT, Adam; COUSTATY, Mickael; DOUCET, Antoine. **Survey of post-OCR processing approaches**. ACM Computing Surveys, v. 54, n. 6, artigo 124, p. 1–37, 2021. DOI: 10.1145/3453476. Disponível em: <https://doi.org/10.1145/3453476>. Acesso em: 30 jul. 2026.

PRODANOV, Cleber Cristiano; FREITAS, Ernani Cesar de. **Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico**. 2. ed. Novo Hamburgo: Feevale, 2013.

SCHOUERI, Luís Eduardo. **Direito Tributário**. 9. ed. São Paulo: Saraiva Educação, 2019.

SONKAR, Shashank; WANG, Zichao; BARANIUK, Richard. MANER: **Mask Augmented Named Entity Recognition for Extreme Low-Resource Languages**. In *Proceedings of the Fourth Workshop on Simple and Efficient Natural Language Processing (SustainLP)*, 2023. p. 219–226, Toronto, Canada (Hybrid). Association for Computational Linguistics.