

16º Congresso de Inovação, Ciência e Tecnologia do IFSP - 2025

Web Crawling para Coleta de Dados de Fundos Imobiliários da B3

Vitória Regina Pereira Santos¹, Gustavo Bartz Guedes¹

¹ Instituto Federal de Educação, Ciência e Tecnologia de São Paulo - Campus Hortolândia

Área de conhecimento (Tabela CNPq): 1.03.03.04-9 Sistemas de Informação.

RESUMO: Este artigo, parte de um projeto de iniciação científica, tem como objetivo apresentar o desenvolvimento de um *crawler* para a coleta automatizada de dados de Fundos de Investimento Imobiliário (FIIs) a partir de fontes como o site Fundos.NET e os Dados Abertos da Comissão de Valores Mobiliários (CVM). O *crawler*, parcialmente implementado, foi desenvolvido na linguagem Python utilizando Scrapy e Selenium. As etapas iniciais envolveram testes de extração de dados e análise estrutural dos sites-alvo. Como resultado parcial, o *crawler* extrai dados básicos dos FIIs e os armazena em um banco de dados relacional. O trabalho tem como objetivo final automatizar a coleta, organizar e consolidar os dados dos FIIs, promovendo eficiência na análise de indicadores.

PALAVRAS-CHAVE: WEB SCRAPING; FUNDOS IMOBILIÁRIOS; AUTOMAÇÃO; DADOS FINANCEIROS; PYTHON; BANCO DE DADOS.

Web Crawling for Data Collection on Real Estate Investment Trusts Listed on B3

ABSTRACT: This paper, part of a scientific initiation project, aims to present the development of a crawler for the automated collection of data on Real Estate Investment Trusts (REITS) from sources such as the Fundos.NET website and the Open Data platform of the Brazilian Securities and Exchange Commission (CVM). The crawler, partially implemented, was developed in Python using Scrapy and Selenium. The initial stages involved data extraction tests and structural analysis of the target websites. As a partial result, the crawler extracts basic data from REITs and stores them in a relational database. The ultimate goal of this work is to automate the collection, organization, and consolidation of REIT data, thereby enhancing the efficiency of indicator analysis. **KEYWORDS:**

WEB SCRAPING; REAL ESTATE INVESTMENT TRUSTS; AUTOMATION; FINANCIAL DATA; PYTHON; DATABASE.

INTRODUÇÃO

Os Fundos de Investimento Imobiliário (FIIs) são um tipo de investimento coletivo em que os investidores aplicam recursos em um fundo, pela compra de cotas, que por sua vez são utilizadas para adquirir e administrar ativos do setor imobiliário (B3, 2024). Esses ativos podem ser imóveis físicos, como shoppings, lajes corporativas, galpões logísticos ou papéis do mercado imobiliário, como

Certificados de Recebíveis Imobiliários (CRI) ou Certificado de Recebíveis do Agronegócio (CRA). As cotas dos FIIs são negociadas nos pregões da Bolsa de Valores como ativos. Atualmente, o mercado de FIIs possui mais de 2,7 milhões de investidores, sendo a maioria pessoas físicas, e movimentada, em média, R\$ 285 milhões por dia (CARVALHO, 2024; Clube FII, 2024).

Os FIIs têm se destacado no mercado financeiro brasileiro, não apenas pelo papel relevante na diversificação de portfólios, mas também pelo seu potencial de geração de renda recorrente. Os FIIs distribuem aos cotistas proventos, que correspondem à parcela do lucro obtido pelo fundo proveniente da exploração de seus imóveis, tais como aluguel ou ganho de capital em operações de compra e venda de ativos imobiliários. Esses proventos são geralmente pagos mensalmente, diretamente na conta do investidor, proporcionalmente à quantidade de cotas possuídas.

Uma das principais vantagens dos FIIs é que os proventos distribuídos aos investidores são, em regra, isentos de Imposto de Renda (IR) para pessoas físicas, desde que atendam aos requisitos legais, como o limite mínimo de cotistas e a negociação em bolsa de valores. Essa característica torna os FIIs uma opção atrativa para investidores que buscam renda passiva estável e eficiência fiscal, potencializando a rentabilidade líquida do investimento.

Os dados dos FIIs são disponibilizados no Fundos.NET, que é uma plataforma, na internet, administrada e mantida pela Bolsa de Valores do Brasil (B3), que centraliza o envio e recebimento de documentos relacionados aos fundos de investimento no Brasil, incluindo FIIs (B3, 2025).

O CVM Dados Abertos é um serviço disponibilizado pela Comissão de Valores Mobiliários (CVM) do Brasil, que permite acesso a dados públicos sobre o mercado financeiro brasileiro. Por meio desse serviço, a CVM divulga informações em formato estruturado (Comissão de Valores Mobiliários (CVM), 2025).

A grande quantidade de dados disponíveis sobre esses fundos, contudo, ainda carece de padronização e acessibilidade. Portanto, este artigo apresenta o desenvolvimento de um *crawler* para extração, coleta e consolidação dos dados de FIIs.

MATERIAIS E MÉTODOS

Inicialmente foi feita uma análise nas páginas do Fundos.NET para identificação dos padrões de navegação e da estrutura das páginas, que estão no formato HyperText Markup Language (HTML).

No caso do Fundos.NET, os dados de cada FII são apresentados em formato HTML estruturado, em sua maioria no formato de tabelas, o que permite a criação de regras precisas para extração. A Figura 1, referente a um FII específico, mostra que todos os documentos disponíveis estão no formato de tabela.

A Figura 2 apresenta os dados referentes aos proventos de um FII em uma data específica, que estão no formato de tabela HTML. Os proventos referem-se a uma remuneração periódica distribuída aos cotistas de um ativo financeiro.

Quanto às tecnologias utilizadas para desenvolvimento do *crawler* houve o emprego da linguagem de programação Python com as bibliotecas especializadas Scrapy (Scrapy Team, 2024), Selenium e (Selenium Project, 2025).

Os *crawlers* são responsáveis por coletar documentos de diversas origens de dados para que possam ser analisados, indexados, pesquisados e extraídos (IBM, 2025). Esses processos envolvem etapas como a descoberta de informações, a identificação de estruturas disponíveis (por exemplo, nomes de visualizações em bancos de dados ou diretórios de sistemas de arquivos) e, posteriormente, a coleta efetiva dos dados para análise.

Figura 1: Tela com tabela dos documentos de um FII no Fundos.NET

Documentos									
Nome do Fundo	Categoria	Tipo	Espécie	Data de Referência	Data de Entrega	Status	Versão	Modalidade de Envio	Ações
VECTIS DATAGRO CRÉDITO AGRONEGÓCIO - FI NAS CADEIAS PRODUTIVAS DO AGRONEGÓCIO RESP LIMITADA	Relatórios	Relatório Gerencial		31/07/2025	07/08/2025 09:49	Ativo	1	AP	
VECTIS DATAGRO CRÉDITO AGRONEGÓCIO - FI NAS CADEIAS PRODUTIVAS DO AGRONEGÓCIO RESP LIMITADA	Comunicado ao Mercado	Outros Comunicados Não Considerados Fatos Relevantes		31/07/2025	31/07/2025 18:55	Ativo	1	AP	
VECTIS DATAGRO CRÉDITO AGRONEGÓCIO - FI NAS CADEIAS PRODUTIVAS DO AGRONEGÓCIO RESP LIMITADA	Aviso aos Cotistas - Estruturado	Rendimentos e Amortizações		31/07/2025	31/07/2025 17:10	Ativo	1	AP	
VECTIS DATAGRO CRÉDITO AGRONEGÓCIO - FI NAS CADEIAS PRODUTIVAS DO AGRONEGÓCIO RESP LIMITADA	Informes Periódicos	Informe Mensal Estruturado		06/2025	15/07/2025 16:23	Ativo	1	AP	

Figura 2: Tela com os dados de proventos de um FII no Fundos.NET

Nome do Fundo:	VECTIS DATAGRO CRÉDITO AGRONEGÓCIO - FI NAS CADEIAS PRODUTIVAS DO AGRONEGÓCIO RESP LIMITADA	CNPJ do Fundo:	42.502.827/0001-43
Nome do Administrador:	INTRAG DTVM LTDA.	CNPJ do Administrador:	62.418.140/0001-31
Responsável pela Informação:	WALTER HIROAKI WATANABE	Telefone Contato:	011 30726090
Data da Informação:	31/07/2025	Ano:	2025

Código ISIN:	BRVCRACF004	Código de negociação:	VCRA11	Rendimento	Amortização
Ato societário de aprovação (se houver)					
Data-base (último dia de negociação "com" direito ao provento)				31/07/2025	
Valor do provento (R\$/unidade)				0,95	
Data do pagamento				13/08/2025	
Período de referência				JULHO 2025	
Rendimento isento de IR*				Sim	

A persistência dos dados coletados é realizada em banco de dados relacional em um Sistema de Gerenciamento de Banco de Dados (SGBD) PostgreSQL¹, cujo esquema foi previamente estruturado para armazenar dados do nome do fundo, o valor do provento, a data de pagamento e a data-base associada a cada registro. A Figura 3 apresenta o Diagrama Entidade-Relacionamento (DER) do banco e dados. Já a Figura 4 mostra o fluxo do processamento dos dados.

RESULTADOS E DISCUSSÃO

No estágio atual, o *crawler* realiza a coleta de dados básicos dos fundos e faz a inserção no banco de dados. O código está disponível em um repositório no GitHub [https://github.com/VitRegina/IC-FIIS].

Apesar do HTML estruturado, a heterogeneidade das estruturas demandou ajustes contínuos no *crawler* para garantir a extração correta das informações.

Como perspectiva de trabalho futuro, prevê-se a implementação da funcionalidade de *download*

¹https://www.postgresql.org

Figura 3: Diagrama Entidade-Relacionamento do Esquema Conceitual do Banco de Dados.

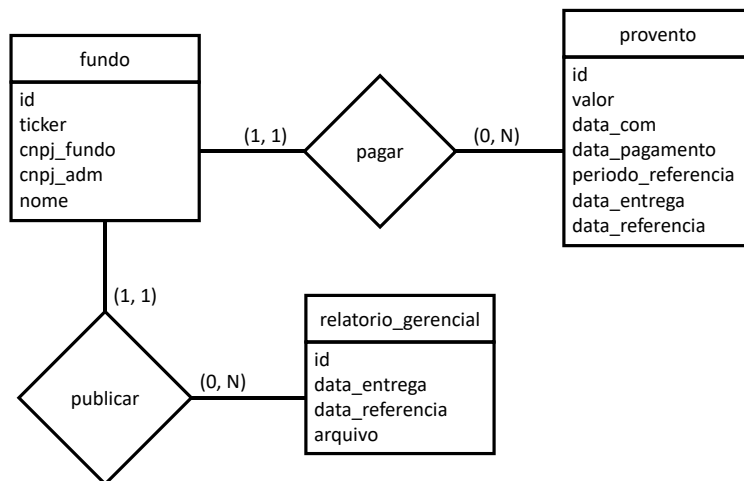
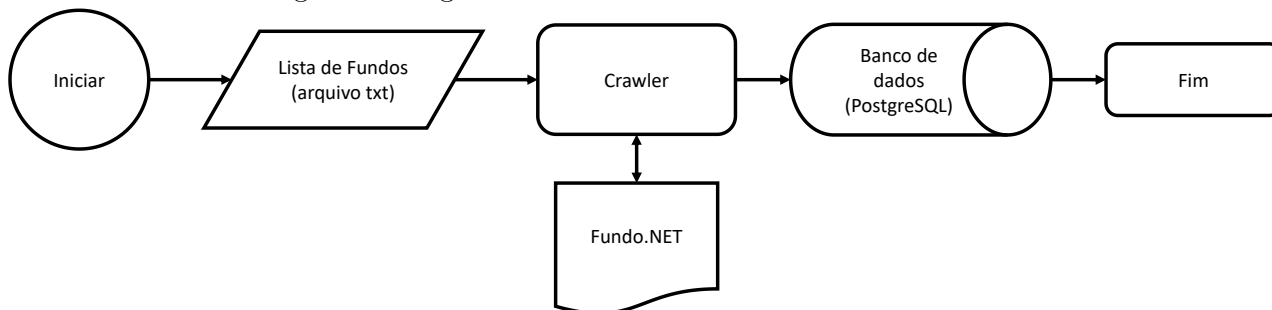


Figura 4: Diagrama de Fluxo de Processamento dos Dados.



dos relatórios gerenciais, que serão armazenados juntamente com seus metadados e organizados em diretórios nomeados de acordo com o código (*ticker*) de cada FII.

Os relatórios gerenciais constituem documentos elaborados e divulgados periodicamente pelas administradoras de Fundos de Investimento Imobiliário (FIIs). Disponíveis em formato PDF, com estrutura não padronizada. Eles contêm informações detalhadas acerca da gestão, desempenho financeiro e operacional do fundo.

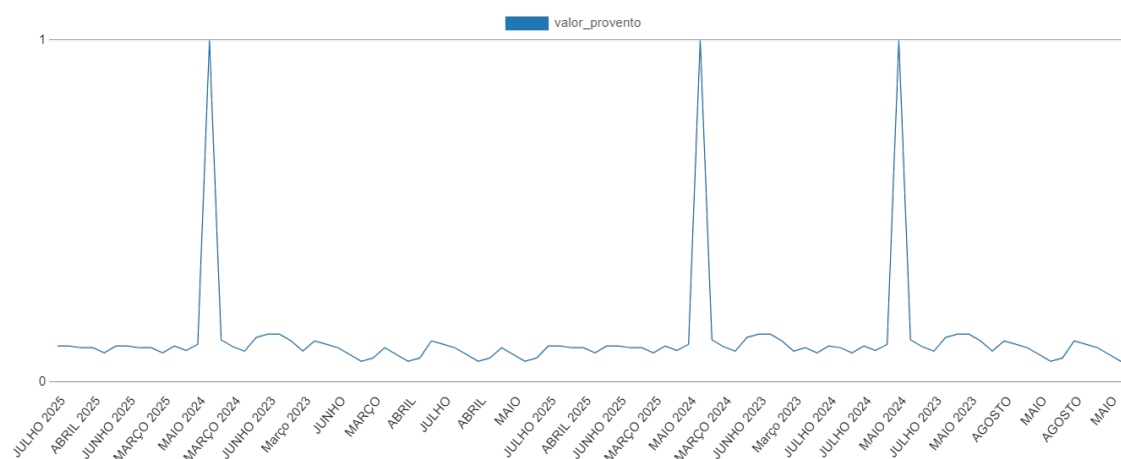
O *crawler* automatizado desenvolvido coletou dados dos fundos imobiliários na plataforma FNET da B3. Durante a execução, foram processadas 7 páginas HTML, totalizando 58 registros de documentos. Cada registro corresponde a um aviso aos cotistas estruturado, sendo que todas as 58 URLs foram acessadas e processadas para extração de dados. Os dados extraídos incluíram: data-base, valor do provento, data de pagamento e período de referência. Todos os registros foram armazenados em um banco de dados PostgreSQL para posterior análise.

Figura 5: Exemplo dos registros armazenados no banco de dados.

Data Output Messages Notifications							
	id [PK] integer	fundo_id integer	valor_provento numeric (10,4)	data_pagamento date	data_base date	periodo_referencia character varying (10)	id_fundo integer
1	270	1	0.1050	2025-08-07	2025-07-31	JULHO 2025	1
2	271	1	0.1050	2025-07-07	2025-06-30	JUNHO 2025	1
3	272	1	0.1000	2025-06-06	2025-05-30	MAIO 2025	1
4	273	1	0.1000	2025-05-08	2025-04-30	ABRIL 2025	1
5	274	1	0.0850	2025-04-07	2025-03-31	MARÇO 2025	1
6	275	1	0.1050	2025-08-07	2025-07-31	JULHO 2025	1
7	276	1	0.1050	2025-07-07	2025-06-30	JUNHO 2025	1
8	277	1	0.1000	2025-06-06	2025-05-30	MAIO 2025	1
9	278	1	0.1000	2025-05-08	2025-04-30	ABRIL 2025	1
10	279	1	0.0850	2025-04-07	2025-03-31	MARÇO 2025	1
11	280	1	0.1050	2024-08-07	2024-07-31	JULHO 2024	1
12	281	1	0.0920	2024-07-05	2024-06-28	JUNHO 2024	1
13	282	1	0.1100	2024-06-07	2024-05-31	MAIO 2024	1
14	283	1	1.0000	2024-06-07	2024-05-31	MAIO 2024	1
15	284	1	0.1230	2024-05-08	2024-04-30	ABRIL 2024	1
16	285	1	0.1030	2024-04-05	2024-03-28	MARÇO 2024	1
17	286	1	0.0900	2023-08-07	2023-08-03	JULHO 2023	1
18	287	1	0.1300	2023-08-07	2023-07-31	JULHO 2023	1
19	288	1	0.1400	2023-07-07	2023-06-30	JUNHO 2023	1
20	289	1	0.1400	2023-06-07	2023-05-31	MAIO 2023	1
21	290	1	0.1200	2023-05-08	2023-04-28	Abril 2023	1

Total rows: 93 of 93 Query complete 00:00:00.302

Figura 6: Evolução do Valor dos Proventos ao longo dos Períodos de Referência



CONCLUSÕES

O desenvolvimento do *crawler* encontra-se em estágio intermediário, mas os resultados obtidos até o momento demonstram a viabilidade técnica da coleta automatizada de dados de FIIs. A utilização das bibliotecas Scrapy e Selenium em conjunto com um banco de dados relacional estruturado evidenciou que é possível organizar e consolidar os dados de maneira sistemática e eficiente.

A principal contribuição deste trabalho consiste em oferecer uma solução que reduz a necessidade de coleta manual de dados, além de consolidá-los em um banco de dados relacional. No caso dos relatórios gerenciais, disponibilizados como arquivo PDF e sem padronização estrutural, a consolidação e organização desses documentos em um repositório estruturado abre caminho para a aplicação de técnicas de Inteligência Artificial (IA), em especial de Processamento de Linguagem Natural (PLN), voltadas à extração de informações atualmente não estruturadas.

CONTRIBUIÇÕES DOS AUTORES

V.R.S. atuou no desenvolvimento do crawler, realização de testes e estruturação do banco de dados.

G.B.G. orientou o trabalho e revisou a escrita final deste artigo.

Todos os autores contribuíram com a revisão do trabalho e aprovaram a versão submetida.

AGRADECIMENTOS

Agradeço ao Programa PIBIFSP pelo apoio técnico e financeiro. Agradeço também ao orientador e demais colegas pelo suporte contínuo durante o projeto.

REFERÊNCIAS

B3. *Fundos de Investimento Imobiliário (FII)*. 2024. <https://www.b3.com.br/pt_br/produtos-e-servicos/negociacao/renda-variavel/fundos-de-investimento-imobiliario-fii.htm>. Acesso em: 30 out. 2024.

B3. *Fundos.NET*. 2025. <<https://fnet.bmfbovespa.com.br/fnet/publico/abrirGerenciadorDocumentosCVM>>. Acesso em: 21 set. 2025.

CARVALHO, W. *Negociação de FIIs em 2024 dispara e bate recorde, aponta B3*. 2024. <<https://www.infomoney.com.br/onde-investir/negociacao-de-fiis-em-2024-dispara-e-bate-recorde-aponta-b3/>>. Acesso em: 30 out. 2024.

Clube FII. *Total de investidores em Fundos Imobiliários atinge 2,71 milhões*. 2024. <<https://www.clubefii.com.br/noticias/Mercado/total-de-investidores-em-fundos-imobiliarios-atinge-271-milhoes>>. Acesso em: 30 out. 2024.

Comissão de Valores Mobiliários (CVM). *CVM Dados Abertos*. 2025. <<https://dados.cvm.gov.br/>>. Acesso em: 21 set. 2025.

IBM. *Crawlers in IBM Content Analytics*. 2025. <<https://www.ibm.com/docs/pt-br/wca/3.5.0?topic=overview-crawlers>>. Acesso em: 24 set. 2025.

Scrapy Team. *Scrapy: framework de raspagem e crawling web rápido e poderoso*. 2024. <<https://www.scrapy.org/>>. Acesso em: 13 ago. 2025.

Selenium Project. *Selenium WebDriver - Documentação oficial*. 2025. <https://www.selenium.dev/pt-br/documentation/webdriver/getting_started/>. Acesso em: 29 jul. 2025.