

UNIVERSIDADE ESTADUAL DE MATO GROSSO DO SUL  
BACHARELADO EM SISTEMAS DE INFORMAÇÃO

Nícolas Herculano Pires

**X Insider: Ferramenta para Extração e Análise de Fundos da B3**

DOURADOS - MS

2024

**Nícolas Herculano Pires**

**X Insider: Ferramenta para Extração e Análise de Fundos da B3**

Projeto Final de Curso I apresentado ao curso de Sistemas de Informação da Universidade Estadual de Mato Grosso do Sul, como requisito parcial para a obtenção do título de Bacharel em Sistemas de Informação.

Orientador: prof. Dr. Ricardo Luís Lachi

DOURADOS – MS

2024

**Nícolas Herculano Pires**

**X Insider: Ferramenta para Extração e Análise de Fundos da B3**

**Banca Examinadora:**

Prof. Dr. Ricardo Luís Lachi (Orientador)

Instituição: Universidade Estadual de Mato Grosso do Sul

Área: Computação

Prof. Dr. Fabrício Sérgio de Paula

Instituição: Universidade Estadual de Mato Grosso do Sul

Área: Computação

Prof. Dr. Nilton César de Paula

Instituição: Universidade Estadual de Mato Grosso do Sul

Área: Computação

DOURADOS – MS

2024

# Resumo

Este trabalho tem como objetivo desenvolver o X-Insider, uma ferramenta que automatiza a coleta e análise de dados sobre fundos de investimentos registrados na B3 (Bolsa de Valores Brasileira). O principal objetivo é proporcionar uma solução eficiente para os usuários, permitindo a extração de informações relevantes e a realização de análises detalhadas, superando as lacunas causadas pelas limitações impostas pela Comissão de Valores Mobiliários (CVM), que restringe o acesso a dados mais recentes no período de 180 dias. O projeto será implementado na linguagem C, utilizando a biblioteca libCurl para realizar requisições HTTP e obter dados da composição das carteiras dos fundos de investimento definidos pelo usuário, com armazenamento dos dados em um banco de dados SQLite. A partir dos dados coletados, serão aplicadas técnicas de extrapolação de dados e resolução de problemas combinatórios, permitindo ao X-Insider identificar padrões, comportamentos históricos e mapear a posição dos fundos. Os usuários poderão realizar consultas automatizadas e visualizar os resultados por meio de uma interface gráfica ou linha de comando. Espera-se que a ferramenta contribua para análises precisas e detalhadas dos fundos de investimento, auxiliando os investidores na tomada de decisões estratégicas.

**Palavras Chaves:** Ferramenta; B3; Decisões estratégicas;

# Sumário

<b>Capítulo 1: Introdução</b> .....	<b>3</b>
<b>Capítulo 2: Revisão da literatura</b> .....	<b>6</b>
2.1. Consulta a dados de fundos de investimentos .....	6
2.2. Linguagem de programação C .....	6
2.3. Biblioteca Curl .....	7
2.4. Banco de dados SQLite.....	7
<b>Capítulo 3: Projeto</b> .....	<b>9</b>
3.1 Definição do conjunto de fundos de investimentos .....	9
3.2 Coleta dos dados do conjunto de fundos de investimentos .....	10
1. Criação de Metadados com libcurl.....	11
2. Coleta de Dados com libcurl.....	15
3. Mapeamento e Relacionamento de Dados .....	15
4. Armazenamento dos Dados Coletados .....	16
3.3 Análise e identificação de padrões e similaridades.....	16
3.4 Cronograma de execução .....	17
Tabela 1: cronograma de execução. ....	18
<b>Referências Bibliográficas</b> .....	<b>19</b>

# Capítulo 1

## Introdução

A diversificação de investimentos é uma prática fundamental no mercado financeiro, destinada a maximizar os ganhos e reduzir os riscos para os investidores. Nesse cenário, os fundos de investimento emergem como uma das alternativas mais populares, sendo amplamente utilizados por investidores que desejam obter retornos superiores por meio da gestão profissional de carteiras diversificadas de ativos. Conforme exposto por Neto (2015), os fundos de investimento operam como "condomínios financeiros", onde investidores (cotistas) aplicam seus recursos coletivamente, buscando retornos a partir da administração de uma carteira de títulos e valores mobiliários.

A relevância dos fundos de investimento é evidenciada pelo papel central que as gestoras de fundos desempenham na alocação de capital. Essas instituições, segundo Bodie, Kane e Marcus (2014), financiam equipes dedicadas à análise e gestão de ativos, com o objetivo de tomar decisões estratégicas informadas sobre a compra e venda de ativos financeiros. No entanto, realizar uma análise cirúrgica do desempenho e das estratégias dos fundos se torna um desafio para investidores individuais, especialmente diante da grande quantidade de dados disponíveis.

Por outro lado, a transparência regulamentada pela Comissão de Valores Mobiliários (CVM) permite o acesso público a dados cruciais sobre fundos de investimento, fornecendo aos investidores informações detalhadas sobre a composição de suas carteiras e o histórico de desempenho. A B3 (Bolsa de Valores Brasileira), por meio de seu portal, disponibiliza esses dados, possibilitando que investidores e pesquisadores façam suas próprias análises e extrações de informações relevantes.

Contudo, essa grande quantidade de dados, aliada às limitações impostas pela própria CVM na consulta desses dados mais recentes, torna necessária a criação de ferramentas capazes de automatizar esse processo e possibilitar análises profundas. Apesar da transparência proporcionada pela regulamentação vigente — onde se exige que fundos de investimento divulguem informações detalhadas de suas carteiras, conforme supervisionado pela Comissão de Valores Mobiliários (CVM) —, a complexidade e o volume de dados disponíveis tornam

difícil para o investidor comum realizar uma análise aprofundada. Adicionalmente, como discutido por Mota (2018), o mercado de capitais está sujeito a práticas abusivas, e a CVM desempenha um papel crucial na repressão dessas práticas para garantir a integridade do mercado. Isso evidencia a importância de ferramentas que automatizam o processo de coleta e análise de dados de fundos de investimento, mitigando assim os riscos de manipulação e otimizando a tomada de decisões.

Com base nesse contexto, o **objetivo geral** deste trabalho será:

- Desenvolver uma ferramenta automatizada, denominada X-Insider, para coleta e análise de dados dos fundos de investimento registrados na B3.

Para atingir esse objetivo geral serão necessários que os seguintes objetivos específicos sejam alcançados:

- Compilar a listagem de fundos de investimentos disponíveis na B3;
- Definir quais as classes dos fundos de investimentos serão utilizados como base para as consultas automatizadas. Por exemplo, considerar somente os fundos de investimentos em ações ou em multimercados;
- Definir se todos os fundos de investimentos das classes escolhidas serão analisados, ou somente um subconjunto deles. Por exemplo, os de gestores mais renomados, os que tem um certo “tempo de vida” no mercado etc.;
- Coletar, no mínimo, os seguintes dados detalhados de todos os fundos de investimentos selecionados: composição da carteira, histórico de investimentos, patrimônio líquido, rentabilidade;
- Analisar os dados detalhados compilados de todos os fundos de investimentos selecionados e identificar padrões e similaridades nos investimentos feitos por eles.

O texto encontra-se organizado do seguinte modo: no **Capítulo 2** é apresentada a revisão bibliográfica, na qual serão detalhadas as ferramentas que serão utilizadas no desenvolvimento deste projeto; no **Capítulo 3**, o projeto a ser desenvolvido é descrito detalhadamente, incluindo o cronograma de execução, metodologia e os resultados esperados; no **Capítulo 4** são apresentadas as conclusões e contribuições deste trabalho.

# Capítulo 2

## Revisão da literatura

Neste capítulo abordaremos a revisão bibliográfica dos principais envolvidos no desenvolvimento da ferramenta X-Insider.

### 2.1. Consulta a dados de fundos de investimentos

A Comissão de Valores Mobiliários (CVM) disponibiliza ao público os dados referentes a fundos de investimentos. Além da CVM<sup>1</sup>, existem ferramentas como Mais Retorno<sup>2</sup> e Status Invest<sup>3</sup>, que também oferecem informações sobre fundos de investimentos. No entanto, ainda que essas plataformas forneçam dados quantitativos, como patrimônio e as ações presentes nas carteiras, não é evidente uma relação entre esses dados, seja para a análise individual de um fundo ou de múltiplos fundos. Este trabalho visa justamente preencher essa lacuna, com o objetivo de obter e correlacionar informações individuais e múltiplas de fundos de investimentos, a fim de gerar dados sólidos e robustos sobre quais ações são efetivamente visadas pelos fundos.

### 2.2. Linguagem de programação C

A linguagem de programação C, desenvolvida por Dennis Ritchie em meados de 1970, tendo seu objetivo principal em entregar uma ferramenta poderosa e versátil no contexto de desenvolver sistemas operacionais. Seu primeiro uso notável foi na reescrita do sistema operacional UNIX, se consolidando como uma linguagem de programação em baixo nível (próximo de comunicação direta como *hardware*), mas com uma sintaxe que facilita a manipulação direta de memória, o que por consequência, torna-se uma escolha comum para criar um sistema de alto desempenho. (Kernighan, Ritchie; 1988). Um dos aspectos mais evidentes da linguagem C é sua portabilidade, o que permite a criação de programas que podem ser compilados em diferentes arquiteturas. Isso é possível principalmente por conta do compilador, o qual traduz a linguagem C para instruções suportadas pela arquitetura corrente. O outro aspecto é referente ao controle explícito da alocação de memória por meio de ponteiros (um recurso que possibilita a administração de endereços de memória), o qual exige um certo amparo técnico, uma vez que o mau uso desta funcionalidade, pode resultar em acessos inválidos ou vazamentos de memória (conhecido também como *overflow*). (Ritchie;

1978)(Kernighan, Ritchie; 1988)

Devido à sua consistência, versatilidade, segurança e alto desempenho, (Ritchie; 1988) decidimos escolhê-la para implementar em nosso projeto.

### **2.3. Biblioteca Curl**

A biblioteca CURL (do inglês, Client for URLs) foi criada por Daniel Stenberg no final dos anos 1990, foi projetada para fornecer uma solução robusta para interações em URLs que vão desde transferências de arquivos até comunicação com APIs de serviços web. (Stenberg; 1997)

Uma de suas características é a simplicidade e versatilidade de implementação, tornando-a adequada para ser utilizada em diferentes Sistemas Operacionais. Além disso, por ser implementada em C, a libCurl (ou biblioteca CURL) tem sua execução leve e eficiente. O motivo da escolha pela CURL se deve a sua abrangência no suporte nos protocolos de requisição, além de controle detalhado de cabeçalhos, proxies e certificados SSL/TLS, o que por sua vez, permite segurança e flexibilidade em requisições web.(Stenberg; 1997).

### **2.4. Banco de dados SQLite**

O SQLite é um sistema de gerenciamento de banco de dados (SGBD) implementado na linguagem C. Desenvolvido em 2000 por Richard Hipp, foi projetado para ser uma solução que não necessita de um servidor separado para operar, o que por sua vez, se torna uma opção no contexto de aplicações que precisam de um armazenamento local, eficiente e de baixo custo.(Gaffney, et.al; 2022).

Uma de suas diferenças de outros SGBDs é não seguir o modelo cliente-servidor, ou seja, os dados são armazenados em um único arquivo de banco de dados, o que simplifica o processo de configuração e manipulação dos dados. Além disso, o SQLite é conhecido por seu desempenho em gravar, ler e recuperar dados da base, conseguindo fazer isso em frações de milissegundos.(Gaffney, et.al; 2022). Por conta dessas qualidades, foi a nossa principal escolha o utilizar para o armazenamento de dados.

No próximo capítulo, será abordado detalhadamente como se procederá este projeto, passando desde a definição dos conjuntos de fundos de investimento, até os objetivos e

cronograma de execução.

# Capítulo 3

## Projeto

Neste projeto será desenvolvida a ferramenta denominada X-Insider. Esse nome é uma analogia ao conceito de “insider” e “X” de incógnita, ou seja, a ferramenta seria como se fosse um agente desconhecido que obtém informações sigilosas de empresas, porém, ao contrário do sentido negativo do termo "insider" no mercado financeiro, tudo o que a ferramenta extrair será com base em dados públicos de forma legal e ética.

Esse desenvolvimento envolve, fundamentalmente, três etapas: 1) definição do conjunto de fundos de investimento que será a base da coleta de dados; 2) a coleta dos dados dos fundos de investimentos selecionados; 3) análise desses dados procurando a identificação de padrões e similaridades nas carteiras de ações desses fundos de investimentos.

### 3.1 Definição do conjunto de fundos de investimentos

A principal fonte de dados para este trabalho será o portal de dados abertos da Comissão de Valores Mobiliários (CVM)<sup>1</sup>. Este portal fornece acesso ao registro histórico<sup>2</sup> da composição da carteira de todos os fundos registrados na B3.

Os dados são disponibilizados ao público em arquivos compactados no formato ZIP, onde cada arquivo é nomeado de acordo com o ano e mês de registro dos fundos, sendo um arquivo para cada mês. (Ver Figura 1)

Para acessar e processar esses dados, o projeto utilizará a ferramenta curl para fazer o download automático dos arquivos contendo as tabelas relevantes do site da CVM. Este método permite uma coleta eficiente e sistemática dos dados necessários para a análise.

1. O processo de extração de dados incluirá as seguintes etapas:
2. Utilização do cURL para acessar as páginas relevantes do portal da CVM.
3. Download dos arquivos ZIP contendo os dados mensais.
4. Descompactação dos arquivos ZIP.

---












<sup>1</sup> <<https://dados.cvm.gov.br>>

<sup>2</sup> <<https://dados.cvm.gov.br/dados/FI/DOC/CDA/DADOS/>>

5. Busca por CNPJs a partir dos arquivos extraídos.
6. Armazenamento dos CNPJs extraídos em um formato adequado para análise posterior.

Para esta abordagem, serão utilizados dados desde 2005 até Setembro de 2024, o que garantirá que tenhamos acesso a um conjunto de dados abrangente e atualizado sobre os fundos de investimento, permitindo uma análise detalhada de suas composições e performance ao longo do tempo.

**Figura 1:** Exemplo de documento de Composição e Diversificação das Aplicações (CDA), detalhando os ativos que compõem as carteiras de fundos de investimento.

 cda_fi_202311.zip	12-Oct-2024 08:23	21M
 cda_fi_202312.zip	12-Oct-2024 08:11	21M
 cda_fi_202401.zip	12-Oct-2024 07:59	21M
 cda_fi_202402.zip	12-Oct-2024 07:47	21M
 cda_fi_202403.zip	12-Oct-2024 07:35	21M
 cda_fi_202404.zip	12-Oct-2024 07:23	20M
 cda_fi_202405.zip	12-Oct-2024 07:09	21M
 cda_fi_202406.zip	12-Oct-2024 06:56	21M
 cda_fi_202407.zip	15-Oct-2024 02:08	15M
 cda_fi_202408.zip	15-Oct-2024 01:53	14M
 cda_fi_202409.zip	15-Oct-2024 01:37	15M

Fonte: feito pelo autor.

### 3.2 Coleta dos dados do conjunto de fundos de investimentos

Com os CNPJs obtidos na etapa anterior, vamos pegar cada um e realizaremos as coletas de dados, a qual será detalhada nas etapas seguintes.

É possível fazer a consulta da composição da carteira de cada um dos fundos com base no CNPJ selecionado, bastando inserir na seguinte URL:

<https://cvmweb.cvm.gov.br/SWB/Sistemas/SCW/CPublica/CConsolFdo/ResultBuscaParticFdo.aspx?CNPJNome=00.000.000/0000-00>

A seguir, passos detalhados de como isso pode ser feito.

## **1. Criação de Metadados com libcurl**

Para ser feita as requisições com a libcurl, foi adotado duas estratégias de se utilizar uma abordagem direta na URL, no caso, temos a primeira que é a estratégia padrão, onde, de acordo com o link padrão da CVM, em que, dividindo em três partes, nós temos:

**URL:** <https://cvmweb.cvm.gov.br>

**Caminho:** /SWB/Sistemas/SCW/CPublica/CConsolFdo/

**Busca por CNPJ:** ResultBuscaParticFdo.aspx?CNPJNome=00.000.000/0000-00

Se juntar todas essas partes, chegamos a seguinte URL:

<https://cvmweb.cvm.gov.br/SWB/Sistemas/SCW/CPublica/CConsolFdo/ResultBuscaParticFdo.aspx?CNPJNome=00.000.000/0000-00>

Na parte “Busca por CNPJ” será inserido o CNPJ e navegado por meio de outras requisições até a composição da carteira. Embora este método já fosse o suficiente, é importante ressaltar as limitações impostas por interação com o site da CVM, em outras palavras, quando inserimos o CNPJ, nós ainda temos requisições pendentes para chegarmos até onde de fato lida com a composição da carteira do fundo de investimento, a seguir é demonstrado os passos a serem seguidos até chegar na composição da carteira:

**Figura 2:** O Site da CVM lista os fundos com base no CNPJ passado como parâmetro

### Consulta Consolidada de Fundo

---

Os seguintes fundos e/ou administradores foram encontrados a partir da sua pesquisa.

Clique sobre o fundo para visualizar seus dados, ou sobre o administrador para visualizar seus fundos associados:

39.573.883/0001-64	KAPITALO TARKUS ADVISORY FUNDO DE INVESTIMENTO EM COTAS DE FUNDO DE INVESTIMENTO EM AÇÕES	FI
--------------------	---	----

[Fale com a CVM](#)

Fonte: feito pelo autor.

**Figura 3:** Opções de visibilidade de dados específicos pela CVM do fundo selecionado na Figura 2.

### Consulta Consolidada de Fundo

Dados Gerais	
<b>Nome do Fundo:</b> KAPITALO TARKUS ADVISORY FUNDO DE INVESTIMENTO EM COTAS DE FUNDO DE INVESTIMENTO EM AÇÕES	<b>CNPJ:</b> 39.573.883/0001-64
<b>Administrador:</b> BTG PACTUAL SERVIÇOS FINANCEIROS S/A DTVM	<b>CNPJ:</b> 59.281.253/0001-23
<b>Situação Atual:</b> EM FUNCIONAMENTO NORMAL	<b>Data Início de Atividades:</b> 21/01/2021
<b>Data de Constituição:</b> 21/01/2021	<b>Web Site:</b> WWW.BTGPACTUAL.COM

Para consultar os regulamentos (a partir de 05/07/2021) e informações cadastrais dos fundos da Instrução CVM Nº 555 e fundos estruturados (FIPs, FILs, FIDCs e FUNCINEs), acesse o sistema SGF: [clique aqui](#)

Documentos Associados
• <a href="#">Balancete</a>
• <a href="#">Composição da Carteira</a>
• <a href="#">Dados diários (Vl. Quota, Patrim. Liq., Num. Cotistas, Captação e Resgate)</a>
• <a href="#">Regulamento</a>
• <a href="#">Documentos Eventuais (assembléias, fato relevante, prospecto distribuição, regulamento, etc)</a>
• <a href="#">Lâmina do Fundo</a>
• <a href="#">Demonstrações contábeis com parecer do auditor</a>
• <a href="#">Perfil Mensal</a>
• <a href="#">Formulário de Informações Complementares</a>

Fonte: feito pelo autor.

**Figura 4:** Ao clicar em “Composição da Carteira”, se obtém uma visão completa da carteira do fundo selecionado.

Patrimônio Líquido do Fundo: R\$ 382.612.614,48			Data de Recebimento das Informações: 11/04/2024 20:06:32							
Lista de Aplicações										
Clique sobre o ativo para mais informações.										
Ativo	Classificação	Empresa Ligada	Negócios Realizados no Mês				Posição Final			
			Vendas		Aquisições		Quant.	Valores		% Patr. Lq.
			Quant.	Valor	Quant.	Valor		Custo	Mercado	
Ações Cod. Ativo: BRF53 Dt. Ini. Vigen.: 10/12/2009	Para negociação	Não	0	0,00	0	0,00	4.740.556		77.365.873,92	20,22
Ações Cod. Ativo: SUZB3 Dt. Ini. Vigen.: 05/08/2004	Para negociação	Não	0	0,00	5.000	311.750,08	921.423		58.952.643,54	15,408
Ações Cod. Ativo: VALE3 Dt. Ini. Vigen.: 24/07/1997	Para negociação	Não	0	0,00	0	0,00	616.716		37.514.834,28	9,805
Ações Cod. Ativo: VBBR3 Dt. Ini. Vigen.: 22/10/2021	Para negociação	Não	0	0,00	0	0,00	996.287		24.917.137,87	6,512
Ações Cod. Ativo: EMBR3 Dt. Ini. Vigen.: 28/07/1997	Para negociação	Não	0	0,00	0	0,00	653.892		21.781.142,52	5,693
Ações Descrição: RRRP3 CNPJ do emissor: 12.091.809/0001-55 Denominação Social do emissor: 3R PETROLEUM OLEO E GAS S.A.	Para negociação	Não	0	0,00	0	0,00	494.964		16.328.862,36	4,268
Ações Descrição: BPAC11 CNPJ do emissor: 30.306.294/0001-45 Denominação Social do emissor: Banco BTG Pactual SA	Para negociação	Sim	0	0,00	0	0,00	304.823		11.104.701,89	2,902
Ações Cod. Ativo: HAPV3 Dt. Ini. Vigen.: 25/04/2018	Para negociação	Não	0	0,00	0	0,00	2.968.159		10.982.188,30	2,87
Ações Cod. Ativo: LREN3 Dt. Ini. Vigen.: 02/01/1998	Para negociação	Não	0	0,00	66.500	1.050.324,01	617.499		10.485.133,02	2,74
Ações Cod. Ativo: ECOR3	Para negociação	Não	0	0,00	0	0,00	901.969		7.351.047,35	1,921

Fonte: feito pelo autor.

Podemos ver que são 3 passos que se deve seguir, o que resulta em 3 interações com o site, que por sua vez, resulta em 3 requisições que devem ser feitas. Ainda que nesses 3 passos não tivesse a complicação filtragem de conteúdo do HTML em C, teríamos ainda o atraso na requisição da CVM. Em outras palavras, se tivéssemos em um cenário hipotético otimista, onde cada requisição levasse 300 milissegundos, no total teríamos 900 milissegundos, apenas para chegar na Composição da carteira, o que poderia ser reduzido em 66.7% se tivéssemos como acessar diretamente a página de composição da carteira. Desta maneira, chegamos à segunda estratégia.

A segunda estratégia, se trata de uma busca direta por uma chave primária única e, teoricamente, permanente. Ao ser analisado a aba de composição da carteira de um fundo na CVM, se deparamos com o CNPJ sendo trocado pelo parâmetro “PK\_PARTIC”, resultando em uma URL que se dividirmos em três partes, chegamos à seguinte maneira:

**URL:** <https://cvmweb.cvm.gov.br>

**Caminho:** /SWB/Sistemas/SCW/CPublica/CDA/

**Busca por Chave Primária:** CPublicaCDA.aspx?PK\_PARTIC=000001

Se juntarmos todas as partes, chegamos na seguinte URL:

[https://cvmweb.cvm.gov.br/SWB/Sistemas/SCW/CPublica/CDA/CPublicaCDA.aspx?PK\\_PARTIC=000001](https://cvmweb.cvm.gov.br/SWB/Sistemas/SCW/CPublica/CDA/CPublicaCDA.aspx?PK_PARTIC=000001)

Após um algoritmo de força bruta baseado em decimais, chegamos à conclusão que a maior “PK\_PARTIC” até o dia 13/10/2024 é 250404, sendo que quando colocado um número maior, o servidor retorna um erro de processamento do servidor, quando menor, tem dois destinos, ou o fundo não tem CDA (Composição de Diversificação das Ações) e o servidor retorna o erro que o fundo não possui o CDA, ou retorna o CDA do fundo pesquisado. Após ter sido monitorado dois CNPJs específicos durante o período do mês 06/2024 até o mês 10/2024, foi constatado que a chave PK\_PARTIC se demonstrou consistente, não mudando o CNPJ vinculado à PK\_PARTIC. Isso por si só, não garante se o CNPJ sempre será estático, por conta disso, será feito um metadata que relaciona CNPJ com PK\_PARTIC, o qual será constantemente atualizado.

## 2. Coleta de Dados com libcurl

Com o CNPJ em mãos, será feita várias requisições, sendo a primeira que será para captar o formato que está os dados e as demais para abranger as datas necessárias para análise.

A primeira requisição será via URL, também conhecida como GET, em que será para captura do HTML e terá um papel importante na identificação de datas.

Já as demais requisições, será feita por um método chamado POST, onde se altera os dados do HTML não se limitando somente à URL

Ambos os métodos têm a proposta de coletar dados, com a diferença que o segundo é para navegar entre as datas e o primeiro é para colher a navegação de datas. Os dados coletados com a libcurl são entregues em um ponteiro para uma cadeia de caracteres (char \*), o qual representa o HTML retornado do servidor na requisição.

## 3. Mapeamento e Relacionamento de Dados

Com os Dados coletados, será feito um mapeamento, onde se filtra do conteúdo as ações componentes da carteira e opções de compra e venda, que estão contidos no HTML, desconsiderando outras informações como por exemplo, aplicações em debêntures. Estes conteúdos serão separados em uma estrutura de dados em C (struct). Não optamos por analisar diretamente após a coleta, pois dificultaria a manutenção aplicação do código em outros contextos.

#### **4. Armazenamento dos Dados Coletados**

Após a separação de dados de maior relevância, os mesmos serão guardados em uma base de dados SQL utilizando a biblioteca SQLite. Dessa maneira, será possível recuperar os dados mesmo após o fechamento do programa, ou até mesmo a inserção de dados por meio de um SGBD sem interação direta com o módulo de requisição HTTP. A escolha do SQLite se deve não somente ao desempenho, mas também à sua comunicação com a linguagem C, a qual possui uma compatibilidade direta com a mesma. Além disso, salvar os dados em SQL torna a administração de dados em um nível elevado, o que facilita a interação do usuário com o programa.

### **3.3 Análise e identificação de padrões e similaridades**

A partir dos dados coletados sobre as carteiras de ações do conjunto de fundos de investimentos selecionados serão analisados e identificados, no mínimo, os seguintes padrões e similaridades nos investimentos feitos por eles:

- Identificação dos ativos presentes de forma significativa na maioria dos fundos de investimentos analisados;
- Determinação de um valor médio de aquisição dos ativos mantidos nas carteiras dos fundos;
- Avaliação da quantidade e do valor financeiro dos ativos mantidos por um fundo de investimento específico;
- Estudo do intervalo de tempo no qual o fundo de investimento realizou as aquisições de ativos em questão;
- Avaliação da movimentação do fundo no quesito aumento ou redução na posição de ativos de um mês para o outro;
- Análise do tempo necessário para que o fundo complete seu posicionamento em um determinado ativo, verificando se isso ocorre em um ou mais meses.

Para ser possível não somente identificar estes padrões, mas também criar o X-Insider, foi proposto o cronograma de execução que será apresentado no item seguinte.

### 3.4 Cronograma de execução

Visando o desenvolvimento do projeto, estão previstas as seguintes atividades:

1. Definição da Proposta;
2. Ajustes Finais e Entrega da Proposta;
3. Apresentação da Proposta;
4. Revisão da libCurl em C;
5. Revisão do SQLite em C;
6. Ferramenta para descobrir o limite aceitável pela CVM;
7. Módulo de Metadados com libcurl;
8. Módulo de Coleta de Dados com libCurl;
9. Módulo de Mapeamento e Relacionamento de Dados;
10. Módulo de Armazenamento dos dados coletados em SQLite;
11. Módulo de Análise de Dados que realiza a análise dos dados coletados;
12. Módulo de Interface trata a exibição dos dados coletados e analisados;
13. Testes de funcionamento e adaptação aos novos módulos.
14. Escrita da Monografia;
15. Análise e Documentação dos resultados obtidos;
16. Ajustes Finais e Entrega do Projeto;
17. Defesa;

Ano	2024								2025											
	Etapas	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov
1	✓	✓	✓	✓	✓															
2					✓	✓														
3						x	x													
4							x	x	x											
5							x	x	x											
6									x											
7										x	x	x								
8											x	x	x							
9												x	x	x						
10														x						
11														x	x	x	x			
12																	x	x		
13											x	x	x	x	x	x	x	x		
14											x	x	x	x	x	x	x	x		
15															x	x	x	x		
16																	x	x		
17																				x

**Tabela 1:** cronograma de execução.

## Referências Bibliográficas

B3 - **Brasil, Bolsa, Balcão**. Disponível em: [https://www.b3.com.br/pt\\_br/b3/institucional/quem-somos/](https://www.b3.com.br/pt_br/b3/institucional/quem-somos/). Acesso em: 13 out. 2024.

Bodie, Zvi; Kane, Alex; Marcus, Alan J. **Investimentos**. 10ª ed. AMGH Editora, 2014. ISBN 8580554209.

Comissão de Valores Mobiliários. **CVM**. Disponível em: <https://www.gov.br/cvm/pt-br>. Acesso em: 13 out. 2024.

Comissão de Valores Mobiliários (CVM). **Resolução nº 172, de 13 de novembro de 2022**. Disponível em: <https://conteudo.cvm.gov.br/export/sites/cvm/legislacao/resolucoes/anexos/100/resol172.pdf>. Acesso em: 13 out. 2024.

Gaffney, Kevin & Prammer, Martin & Brasfield, Larry & Hipp, D. & Kennedy, Dan & Patel, Jignesh. (2022). **SQLite: past, present, and future**. Proceedings of the VLDB Endowment. 15. 3535-3547. 10.14778/3554821.3554842.

IPEA - Instituto de Pesquisa Econômica Aplicada. **Relatórios e publicações**. Disponível em: <https://repositorio.ipea.gov.br/handle/11058/6489>. Acesso em: 13 out. 2024.

Kernighan, Brian; Ritchie, Dennis. **The C Programming Language**. 2ª ed. Prentice Hall, 1988. Disponível em: <https://github.com/auspbro/ebook-c/blob/master/The.C.Programming.Language.2Nd.Ed%20Prentice.Hall.Brian.W.Kernighan.and.Dennis.M.Ritchie..pdf>. Acesso em: 13 out. 2024.

Mota, Fernando de Andrade. **Abusos no mercado de capitais: conceituação e repressão pela Comissão de Valores Mobiliários**. 2018. Tese (Doutorado) — Universidade de São Paulo.

Neto, Alexandre Assaf. **Mercado Financeiro**. 4ª ed. Amazon, 2018. Disponível em: <https://www.amazon.com.br/Mercado-Financeiro-Alexandre-Assaf-Neto/dp/8597001933>. Acesso em: 13 out. 2024.

Ritchie, Dennis. **The Development of the C Language**. Bell Labs, 1978. Disponível em: <https://www.bell-labs.com/usr/dmr/www/chist.pdf>. Acesso em: 13 out. 2024.

Stenberg, Daniel. **Curl: A Command Line Tool and Library for Transferring Data with URLs**. 1997.

XP Investimentos. **Fundos de ações ativos podem esconder a carteira por até 6 meses**. 2024. Disponível em: <https://conteudos.xpi.com.br/fundos-de-investimento/relatorios/cvm->

[fundos-de-acoes-ativos-podem-esconder-a-carteira-por-ate-6-meses/](#). Acesso em: 13 out. 2024.