IMT 542 Spring 2025

Selina Liu & Daisy Deng

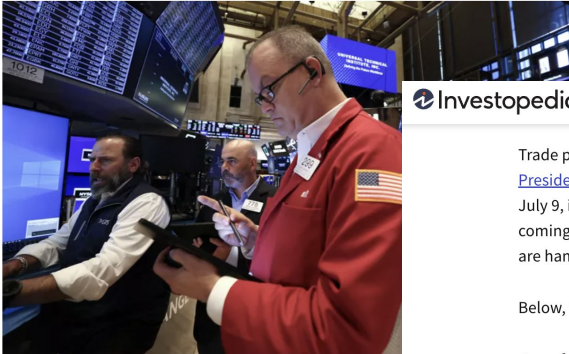# Finance Definition Extractor

AGENDA

INFORMATION STORY

Problem

- Financial terms are hard to understand, especially in unstructured text

- No fast or easy way to get definitions in context

- Existing sites like Investopedia are built for reading, not quick lookup or automation

These Are Stocks You Should Watch in June

By COLIN LAIDLEY · Published May 31, 2025 07:00 AM EDT

May was the best month for the S&P 500 and Nasdaq Composite since November 2
Credit: Timothy A. Clary / AFP / Getty Images

Stocks soared in May, lifted by easing trade tensions betw
China, a strong end to first-quarter earnings season, and
remains on solid footing. The S&P 500 and Nasdaq Comp
their best months since November 2023, rising 6.2% and

Trade policy is likely to remain Wall Street's primary focu
President Trump's "Liberation Day" tariffs expected to g

Trade policy is likely to remain Wall Street's primary focus in June. With President Trump's "Liberation Day" tariffs expected to go back into effect on July 9, investors will be hoping the White House strikes more trade deals in the coming weeks. Wall Street will also be watching Congress, where Republicans are hammering out the details of Trump's 'One Big, Beautiful Bill.'

Below, we look at a few stocks to keep an eye on this month.

Apple

Tariffs have been the primary focus of Apple (AAPL) investors in recent months, but their attention is likely to shift to artificial intelligence when the company hosts its Worldwide Developer Conference (WWDC) on June 9.

Last year's WWDC saw the unveiling of Apple Intelligence, the company's proprietary artificial intelligence offering. Executives touted Apple Intelligence's personalization and privacy features, and showed off a few AI applications like image and emoji generators.

At this year's WWDC, Apple is reportedly planning to release a software development kit that enables third parties to build features using the large language models underpinning Apple Intelligence. [1] Apple has disappointed Wall Street and some users with its slow AI roll-out. Opening up Apple Intelligence to outside parties could satisfy the critics by accelerating the development of AI apps for the iPhone and other AI-enabled devices.

INFORMATION STORY

Target Audience

- Students and new investors
- Self-learners and educators needing quick references

What Our Tool Does

- Users paste a paragraph (≤100 words)
- Tool extracts and defines financial terms using Investopedia
- Returns structured JSON: term, definition, and URL
- Saves time and supports learning in context

IN SCOPE

- **Input:** Free-form user text (up to 100 words)
- **Output:** Financial term definitions (if available on Investopedia)
- **Structure:** JSON with term, definition, URL
- **Access:** REST API (PUT /input, GET /definitions)

OUT OF SCOPE

- PDF/CSV inputs
- Definitions beyond Investopedia scope
- Non-English terms

EXISTING INFORMATION STRUCTURE

Source: Investopedia.com

Original Format

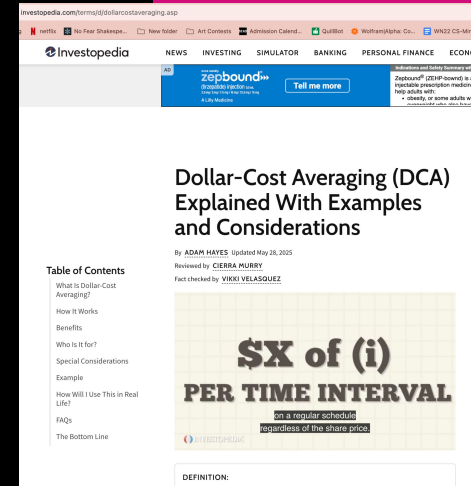-   Articles hosted on static URLs like: https://www.investopedia.com/terms/b/bearmarket.asp
-   Introductory paragraph(s)
-   Linked related terms
-   Inconsistent use of HTML IDs/classes

Access Method

-   No official API
-   Manual or automated access via HTML scraping (using Python + BeautifulSoup)

Content Characteristics

-   Content is written for human reading, not machine parsing
-   **<p id="mntl-sc-block_2-0">** (long-form definition)
-   **<div id="short-definition__text_1-0">** (short definition)
-   Embedded links point to related Investopedia articles

FAIR ASSESSMENT

F
FINDABLE

A
ACCESSIBLE

I
INTEROPERABLE

R
REUSABLE

Definitions are matched using precise term extraction.

Each result links back to the original Investopedia URL.

Public REST API with simple endpoints.

Lightweight JSON output ensures compatibility across platforms.

Output is structured JSON with flat key-value pairs.

Usable in websites, dashboards, educational apps, and more.

Clean formatting enables copy-paste into documents or apps.

External tools can easily consume the API response.

TRANSFORMATION GOALS

Current Deficiency

→ Unstructured HTML across pages

→ No API or structured data format

→ Manual search for each financial term

→ No batch lookup — one term at a time

→ No term-level structure — just full pages

→ Content not interoperable

What We Plan to Do

→ Scrape elements using tag IDs

→ Convert scraped content into clean, portable JSON format

→ Accept full paragraphs and auto-extract valid terms using NLP tokenization

→ Extract all relevant terms from one user input block

→ Return lightweight objects with just term, definition, and url

→ Build REST API

New Structure

→ Reliable and consistent definition extraction

→ Easily reusable by any app, dashboard, or interface

→ Resilient to URL mismatches

→ Minimal, structured output optimized for UX, data reuse, and analysis

→ Can be integrated into websites, chatbots, BI tools, and education platforms

NEW PORTABLE INFORMATION STRUCTURE

- **Format:** JSON
- **Structure:** Flat key-value pairs
- **Portable across apps:** web, mobile, dashboard, chatbot
- **Access Method:** Lightweight REST API (GET, PUT)

```
1   [
2     {
3       "term": "return on equity",
4       "definition": "Return on equity (ROE) is a measure of a company's financial
        performance. It is calculated by dividingnet incomeby shareholders' equity.
        Because shareholders' equity is equal to a company\u2019s assets minus its debt,
        ROE is a way of showing a company'sreturn on net assets.",
5       "url": "https://www.investopedia.com/terms/r/returnonequity.asp"
6     },
7     {
8       "term": "dividend yield",
9       "definition": "A stock's dividend yield is a ratio showing how much a company
        pays out in dividends each year relative to its stock price. The reciprocal of
        the dividend yield is the dividend payout ratio.",
10      "url": "https://www.investopedia.com/terms/d/dividendyield.asp"
11    },
12    {
13      "term": "bear market",
14      "definition": "A bear market is a market condition where investors are more
        risk-averse than risk-seeking, defined by some as when prices have fallen more
        than 20% from previous highs.",
15      "url": "https://www.investopedia.com/terms/b/bearmarket.asp"
16    },
17    {
18      "term": "liquidity",
19      "definition": "Liquidity refers to the efficiency or ease with which
        anassetorsecuritycan be converted into ready cash without affecting
        its\u00a0market price. The most liquid asset of all is cash itself. Consequently,
        the availability of cash to make such conversions is the biggest influence on
        whether a market can move efficiently.",
20      "url": "https://www.investopedia.com/terms/l/liquidity.asp"
21    },
22    {
23      "term": "dividend",
24      "definition": "Dividends are a percentage of a company's earnings paid to its
        shareholders as their share of the profits. Dividends are generally paid
        quarterly, with the amount decided by the board of directors based on the
        company's most recent earnings.",
25      "url": "https://www.investopedia.com/terms/d/dividend.asp"
26    },
```

React Frontend

Hosted on:
https://selinaliu8.github.io/imt542

→ Built using **create-react-app**

→ Paste or type up to 100 words

→ Click a button to analyze the paragraph

→ Triggers a PUT request to the backend's /input route

→ Then issues a GET request to /definitions to retrieve matched results

→ Clean rendering of definitions in cards with clickable links

Flask Backend

Hosted on:
https://financedefinitionextractor-<...>.azurewebsites.net

→ Built with **Flask + BeautifulSoup + requests**

→ PUT /input: Accepts raw text and stores it in input.txt

→ GET /definitions: Parses the saved text, extracts candidate terms, and returns matching Investopedia definitions

→ Bigram and trigram term detection

→ HTML parsing from known Investopedia structures

→ JSON formatting of results

→ Supports both short-form and full-article definitions

Investopedia Definition Finder

Paste up to 100 words here...

Get Definitions

Investopedia Definition Finder

The stock market, a complex ecosystem of buyers and sellers, can be both exhilarating and daunting. Investors, from seasoned professionals to first-timers, navigate a landscape of fluctuating prices, quarterly reports, and macroeconomic trends. Personal finance, the bedrock of individual well-being, involves budgeting, saving, and investing wisely.

Get Definitions

Found Definitions:

personal finance: Personal finance is a term that covers managing your money as well as saving and investing. It encompasses budgeting, banking, insurance, mortgages, investments, retirement, tax, and estate planning. The term often refers to the entire industry that provides financial services to individuals and households and advises them about financial and investment opportunities.
Read more

stock market: The stock market is a broad term for the network of exchanges, brokerages, and over-the-counter venues where investors buy and sell shares in publicly traded companies. Though people sometimes use "stock market" to refer to theNew York Stock Exchange(NYSE) or theNasdaq, these exchanges are components of a wider global marketplace.
Read more

investing: Investors allocate money or capital to work on a project or undertaking to generate positive returns, or profits. The type of gains depends on the project or asset. Real estate investing can produce both rents and capital gains. Many stocks pay quarterlydividends. Bonds commonly pay regular interest.
Read more

finance: Finance is a broad term that encompasses managing and investing money to preserve it, increase it, or direct its use. Individuals, companies, and governments all engage in finance when they put their money to work.
Read more

market: A market is any place or venue where buyers and sellers can exchange goods and services. A market may be physical, like a retail outlet, or virtual, like an online brokerage with no physical contact between buyers and sellers.
Read more

[Github Repo](#)

DEMO

(Try it out yourself!)

Quality & Performance

01   Term Accuracy   Ensure extracted terms match real Investopedia entries.

02   Definition Clarity   Validate that definitions are clear and easy to read.

03   API Response Time   Monitor response time and error rates.

04   Resilience   Use tag-based scraping for stable parsing.

Future Development Plan

Optimize Scraping Efficiency

Goal: Reduce response times to under 5 seconds per request

- Implement request throttling and intelligent caching
-
- Use asynchronous scraping with aiohttp and asyncio to fetch multiple terms in parallel

AI-Based Text Tokenization

Goal: Improve matching accuracy for multi-word financial terms

- Use an AI API to intelligently extract key financial terms and phrases from user input
- Send user text to the API and receive a structured list of the most relevant finance-related entities for definition lookup

Chrome Extension Integration

Goal: Extend definition-finding capability to any website

- Develop a lightweight Chrome extension that allows users to highlight text and instantly fetch Investopedia definitions
- Communicates with our backend API for consistent result formatting

MSIM Intersection

Information Architecture

Used structured JSON to organize financial terms clearly.

Data Science

Applied tokenization and bigram/trigram detection to extract terms.

UX Design

Built a simple, user-friendly interface for easy access.

Conclusion

- Our project makes financial knowledge accessible via automation and clean UI.

- By scraping and structuring Investopedia content, we empower users to learn independently.

- The tool is lightweight, flexible, and ready for use in education, fintech, and research.

- This bridges a gap between complex finance content and user-friendly tools.

THANK YOU

Any questions?