International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868
Foundation of Computer Science FCS, New York, USA
Volume 12– No.45, October 2024– www.ijais.org

Beyond Keywords: Understanding User Intent for Personalized Financial Search

Siddharth Dabhade, PhD
Assistant Professor,
School of Management Studies,
National Forensic Sciences University,
Gandhinagar, Gujarat

Prayas Gajbhiye
UG Research Scholar,
School of Management Studies,
National Forensic Sciences University,
Gandhinagar, Gujarat

Himanshu Thakkar, PhD
Assistant Professor,
School of Management Studies,
National Forensic Sciences University,
Gandhinagar, Gujarat

ABSTRACT
Traditional financial information retrieval systems primarily rely on market data feeds and predefined search options. While these systems offer a diverse overview of financial information, traditional search engines may not always fully understand the specific information needs behind a user's search. This study explores an elegant approach that powers accurate user query analysis to personalize search results and cater to the exact information goals of financial users. The proposed model analyses user queries to understand the intent behind a search. This allows users to get results and recommendations, providing a more nuanced and user-centric experience than traditional approaches. The researchers discuss this approach's potential benefits, including improved user experience, understanding of user intent, and the ability to anticipate user needs. This study also listed the challenges associated with user data collection and privacy concerns. This research contributes to developing comprehensive, user-centric, and personalized next-generation financial information retrieval systems.

Keywords
Domain-Specific Search Engine, User Query Analysis, Financial Search, Information Retrieval, Personalization, User Intent, User-Centric Search, Entity Type Filtering and Information Retrieval in Finance.

1. INTRODUCTION
Artificial intelligence is more vast than human expectation. It is also the constantly evolving connecting field of ML (Machine learning) and computer science. Artificial intelligence (AI) can perform tasks as humans can, even better than humans in some areas, cost-saving, and it is programmed to work continuously without the need for rest, increasing productivity and efficiency. In the financial sector, AI has emerged as a transforming tool that states how economic data is collected, analyzed, and used for decision-making. It makes repetitive tasks more straightforward (Odonkor et al., 2024).

The accounting field has recently shifted substantially, especially regarding the introduction of AI. Artificial intelligence technological advancement constantly changes traditional accounting techniques, which were previously done manually. This development showcases a fundamental change in financial data management techniques, examination, and reporting (Lark Editorial Team, 2023).

The fast-changing nature of financial markets: Financial professionals and investors need to stay updated on market trends, company news, and regulatory changes to make enhanced decisions.

Information overload: The financial sector generates massive data from diverse sources. It is challenging to find relevant and reliable information quickly.

Time-bound: Financial professionals often have given time boundaries for research. An efficient information retrieval system can significantly improve their productivity and performance.

The financial services industry operates on timely and accurate information. To make clear, concise decisions regarding any financial investment and managing risk, professionals require access to a vast amount of economic data, including market trends, company performance, and regulatory involvement of the company. The continuous increase in the flow of financial information has created a challenge of information overload.

While existing search engines have a significant role in information retrieval, they are not programmed to meet the specific needs of financial or user-intent financial information retrieval. In responses to financial queries, it can be challenging to find relevancy with the user intent, and ranking of the relevant or reliable sources may not be prioritized, which makes the result vaster and makes it more challenging to make decisions faster. Additionally, general search engines do not have the functionalities to refine searches for specific financial relevancy.

In AI, a Domain-Specific Search Engine (DSSE) is a trained LLM (Large Language Model) used to generate and understand the terminology used in a particular field, industry, or domain. To help them understand the industry's domain-specific context, these systems are trained by unsupervised machine learning techniques. Also, this system can solve financial information retrieval challenges. In the context of indexing and ranking the scrawled information within the economic domain, DSSEs prioritize more relevant search results. To offer advanced search features, DSSEs use domain-specific rankings algorithms efficiently, enabling users to find financial data and insights as they need.

DSSEs use Natural language processing (NLP) to identify the intent of user queries and provide them with the most relevant financial information.

2. LITERATURE SURVEY
Traditional information retrieval systems use a "One size fits all" approach. However, the availability of Big Data in finance creates a demand for more user-centric systems. (Allan et al., 2002). The proposed financial DSSE utilizes a user-centric approach by using integrated personalization techniques. By

understanding user profiles, investment goals, and risk tolerance through user interaction, the DSSE can entertain search results and primarily focus on relevant financial information. This user-centric approach promises that financial professionals and investors will be presented with the most specific data for their particular needs, saving them valuable time and effort in solving the complexities of financial information retrieval. General search engines biasedly retrieve the information when a limited number of entities are repeated in search results for various queries. This can restrict the probability of lesser-known entities and hinder various search outcomes. (Penha et al., 2023) If anyone searches for the same entity, it will also get the same results. Here, the user hinders the lesser-known entities; the presented model can help the user get the information he needs. It will rank the websites according to real-world data monitoring. Within the past several years, computers have played an important role in stock prediction. Many of these systems rely on automating existing fundamental and technical strategies to achieve better returns than humans by removing the emotion and bias from trading. (Gidófalvi, 2001) .

The lacking side of these types of systems is that they ignore the news articles, such as losing a costly court battle. To work effectively and more enchantedly, these systems must translate news events into numeric data before appropriate decisions can be made. This problem indicates severe problems in decisions, and in some cases, human analysts must override trades (Schumaker et al., 2012). DSSE can understand the news-related terms and will focus only on the financial data. It will be more efficient to suggest the stock. It will overlook all factors, such as public image news articles, by monitoring real-time data and using it to make predictions.

2.1 AI and NLP

NLP (natural language process) allows machines to regenerate and process natural human-like language. (Wang et al., 2020). Another use of natural language processing is sentiment analysis, a set of processes designed to analyze and monitor the web to determine users' opinions regarding a particular brand, product, or service. This analysis is becoming increasingly essential in the structuring of advanced marketing strategies. In the field of business, NLP algorithms can enhance Document AI applications (Cui et al., 2021) (the work related to extraction of insights and structured data from unstructured contents like multimedia to allow the automation of their operations and classification)

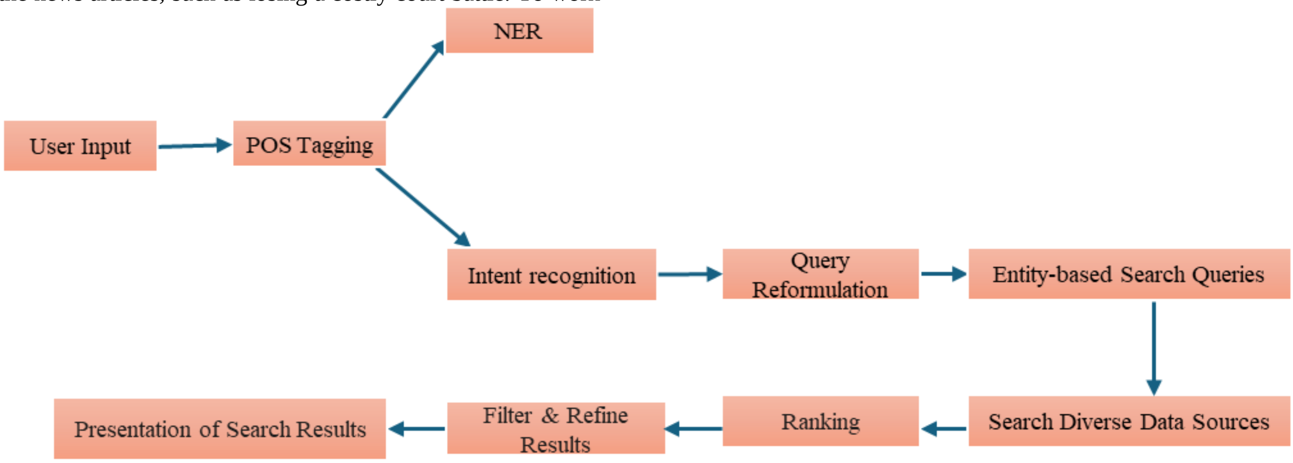

Figure 1. Overview of workings of the algorithm of the proposed model.

(Source: Author's Creation)

3. PROPOSED SYSTEM:

3.1 Understanding User Queries with NLP Techniques

This section explores how the proposed financial DSSE utilizes Natural Language Processing (NLP) techniques to deal with user queries effectively and efficiently. This technique forms the foundation for lower retrieval bias and presenting relevant financial information (see Figure 1).

3.2 Query Parsing and Tokenization:

First, it involves breaking down the user's query into single words or phrases called tokens. This helps for further linguistic analysis.

3.3 Part-of-Speech (POS) Tagging:

NLP also allows a grammatical role (e.g., noun, verb, adjective) to each token within the query. This helps identify the critical components of the query and the user's potential intent.

3.4 Named Entity Recognition (NER):

A core NLP technique is essential for financial information retrieval. NER identifies, understands, and separates specific entities within the query, focusing on the following:

• Financial Instruments: Stocks, bonds, currencies, derivatives (e.g., recent trends for Bitcoin?)

• Companies and Organizations: Publicly traded companies and financial institutions (e.g., How has Apple performed compared to its competitors?)

• Economic Indicators: GDP growth rates, unemployment figures, interest rates (e.g., What would be the impact of inflation on the stock market?)

3.5 Intent Recognition:

By analyzing the query structure, Parts of Speech (POS) tags, and identified entities, the system aims to understand the user's expected result. Examples include:

• Research: Users want information about a specific company, financial instrument, or economic trend.

• Comparison: The user wants to compare the performance of different investments or companies.

• Investment Decision-Making: The user wants to know insights to inform investment choices.

3.6 Query Reformulation (Optional):

Based on the extraction, the system may reformulate the query to get a broader view search and overcome retrieval bias. This could involve Identifying synonyms or alternative terms for financial instruments (e.g., "tech stocks" instead of specific company names) and expanding the timeframe for financial data retrieval, including relevant news articles or analyst reports with traditional financial data sources.

3.7 Searching and Ranking Relevant Financial Information

This section outlines how the proposed DSSE searches and ranks relevant financial information from various sources:

3.8 Diverse Data Sources

The DSSE will access a comprehensive range of financial data sources to minimize reliance on a limited set of entities and overcome retrieval bias. This may include:

Historical and real-time financial data on companies, markets, and instruments (e.g., stock prices, earnings reports, economic indicators).

Financial news archives, analyst reports, and social media data to capture current events, market sentiment, and diverse perspectives. Publicly available filings from companies and regulatory bodies provide in-depth financial information.

3.9 Search Mechanism

The system uses the user's query and the extracted entities (identified through NLP) to generate search queries for every relevant data source. This gives a targeted search that retrieves information specific to the user's financial information needs.

3.10 Ranking Algorithm

A key component for strength retrieval bias is the ranking algorithm taken by the DSSE. This algorithm has many factors to prioritize the most relevant and valuable results for the user. These factors, like relevancy, ensure the minor obstacle with the content that the user needs and what are the things mentioned in their query. Higher weightage to the information from reputable financial institutions established news outlets, or trusted research analysts. Prioritizing recent and up-to-date financial information ensures users gain the latest insights. Ensure a balance between well-known and lesser-known sources for adding value to their decision-making.

3.11 Result Filtering and Refinement (presentation of search results)

The DSSE allows users to refine their search results further using advanced faceted search functionalities. This will enable users to filter based on specific criteria such as Entity Type (Focus on particular companies, instruments, or economic indicators.). Date Range (Narrow down the timeframe for the retrieved financial data.) Source Category (prioritize results from news articles, analyst reports, or regulatory filings based on their information needs), focusing only on the most related and relevant information and prioritizing information or sources based on the “intent of search” and not only the keyword matching and visit counts make result in more valuable and Actionable advice.

4. METHODOLOGY

This subsection details the methods we will employ to gather and prepare the data necessary for developing and evaluating the proposed financial DSSE:

4.1 Financial Data Collection

Structured Financial Data will collect structured financial data from different sources, including financial databases like Bloomberg and Reuters, for historical and real-time financial data on companies, markets, instruments, stock prices, earnings reports, and economic indicators. Regulatory filings SEC Edgar, Ministry of Corporate Affairs (MCA), Reserve Bank of India (RBI), National Stock Exchange of India (NSE) and Bombay Stock Exchange (BSE), Commercial Databases like Thomson Reuters Eikon for publicly available financial disclosures from companies Indian Government Open Data Portal. Unstructured financial data will also be collected to capture diverse perspectives and insights, such as news articles and market commentary from reliable financial newsletters and social media platforms, analyst reports, and research papers from financial institutions and independent research firms.

4.2 Sample Methodology Paragraph Leveraging BloombergGPT

Constructing a robust NLP model for our financial domain-specific search engine (DSSE) will use insights from BloombergGPT, a large language model (LLM) specifically trained on financial text data. While it may not have direct access to all of BloombergGPT's training data, it will analyze publicly available research papers and presentations by Bloomberg researchers. This analysis will focus on the types of financial data (news articles, filings, reports) and entities (companies, instruments, indicators) pointed out during BloombergGPT's training. These insights will inform the model’s data collection strategy, prioritizing similar financial data sources and with the capacity to develop our training dataset to present entities strongly represented in BloombergGPT's training. One of their training data is FinPile (see Figure 2). It comprises many financial data subsets (e.g., news articles and fillings). By strategically utilizing the knowledge obtained from BloombergGPT, this research aims to develop a highly effective NLP model to accurately interpret user queries and retrieve relevant financial information within our DSSE.

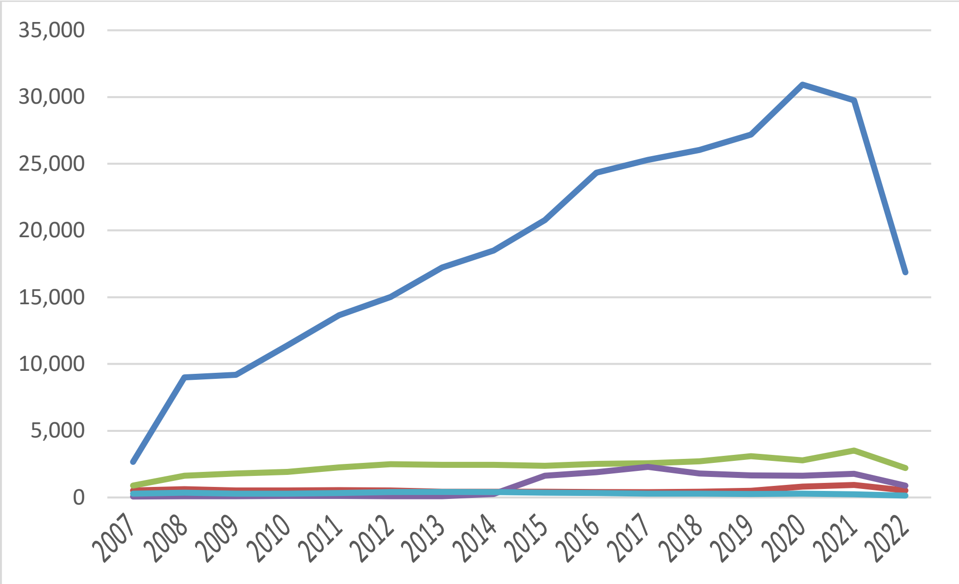

Figure 2: The number of tokens (in millions) contained within documents in FinPile (Y Axis), organized by year (X Axis) and amount (column). Units are millions of tokens.(Wu et al., 2023)

Given (Figure 2) the reported composition of FinPile, our priority is to collect similar financial data sources to improve the training dataset for our NLP model. (Wu et al., 2023). By prioritizing the collection of similar financial data sources for our DSSE, we aimed to create a training dataset that aligns with the potential strengths of BloombergGPT. This included gathering regulatory filings, news articles, and press releases to provide the NLP model with comprehensive and diverse financial information.

5. USER QUERY LOG COLLECTION

Google Finance and Yahoo Finance prioritize predefined search query users enter search terms like ticker symbols, company names, or financial keywords. The platforms do not directly analyze the full user query but focus on identifying relevant keywords within the input. They then match these keywords with data points from real-time and historical market feeds, including stock prices, financial ratios, company news, and analyst ratings. This approach offers a comprehensive overview of financial information while protecting user privacy and allowing platforms to handle massive user traffic. However, it may not always fully collect the specific information needed for a user's search, leading to a less personalized user experience.

For example, one user searches for "Tesla earnings report" on Yahoo Finance. The platform categorizes "Tesla" as a company name and "earnings report" as keywords. It doesn't analyze the entire query as a sentence but focuses on the keywords. Yahoo Finance then queries its market data feeds for recent news articles or press releases from Tesla that have the term "earnings report." If such an article exists, it will catch the attention and be displayed on the search result page. Additionally, you might see the company's stock price and related financial data.

Unlike Google Finance and Yahoo Finance, which primarily focus on market data feeds and predefined searches, the DSSE model takes a user-centric approach by analyzing user queries. This allows the model to understand the demanded information needs and goals the user wants to search for. Where existing platforms offer a comprehensive overview of financial information, the presented model can take user experiences deeper and personalize by tailoring results and recommendations based on the user's specific intent. This focus on user query analysis allows us to provide a more nuanced and relevant search experience than traditional financial information retrieval systems.

For example, a financial advisor is using two different financial information retrieval systems:

System 1: Google Finance (or Yahoo Finance) (Relies on market data feeds and predefined search options)

System 2: (purposed model) User Query-Based Model (Analyzes actual user queries)

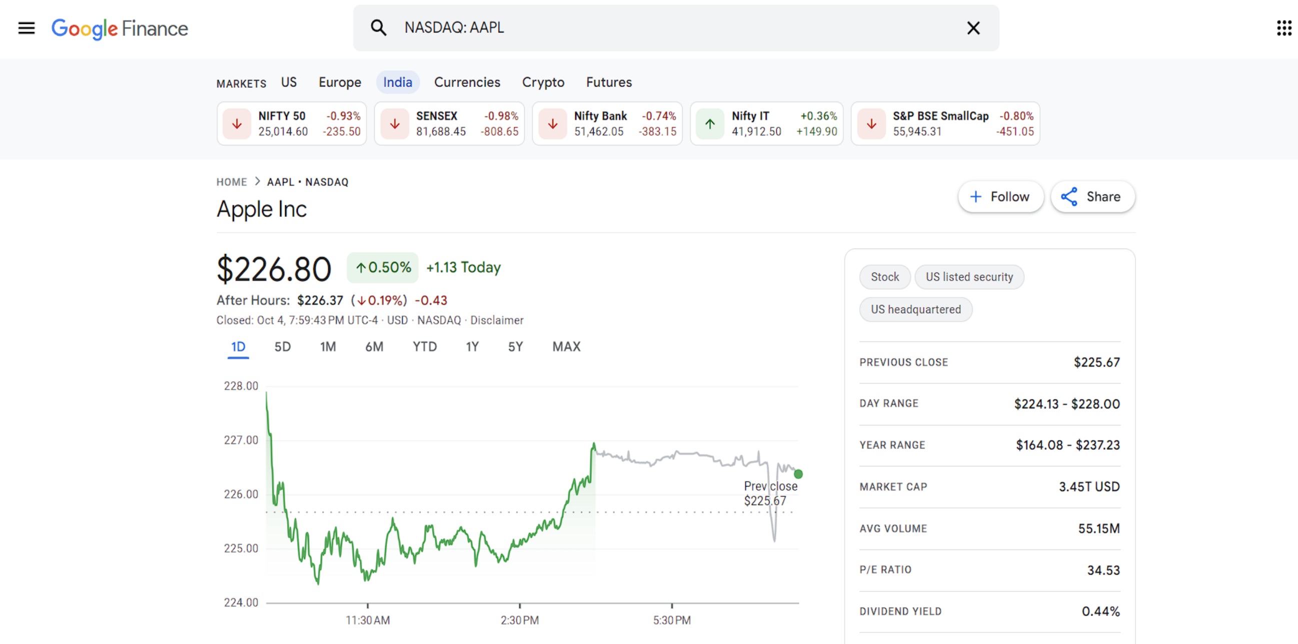

Figure 3. Google Finance search "AAPL" keyword search

(Source: https://www.google.com/finance/?hl=en%20&%20https://finance.yahoo.com/.)

5.1 Scenario 1: Basic Stock Information Retrieval

User on System 1: search for "AAPL" (Apple Inc). The system sees it as a ticker symbol and displays the current stock price, charts, and basic company information (see Figure 3).

User on System 2: search for "AAPL." The system displays the same information as System 1 and analyzes your past queries to see if you're primarily interested in technical analysis charts or fundamental analysis (financial ratios). It also highlights rank relevant data points based on your past behavior/interest.

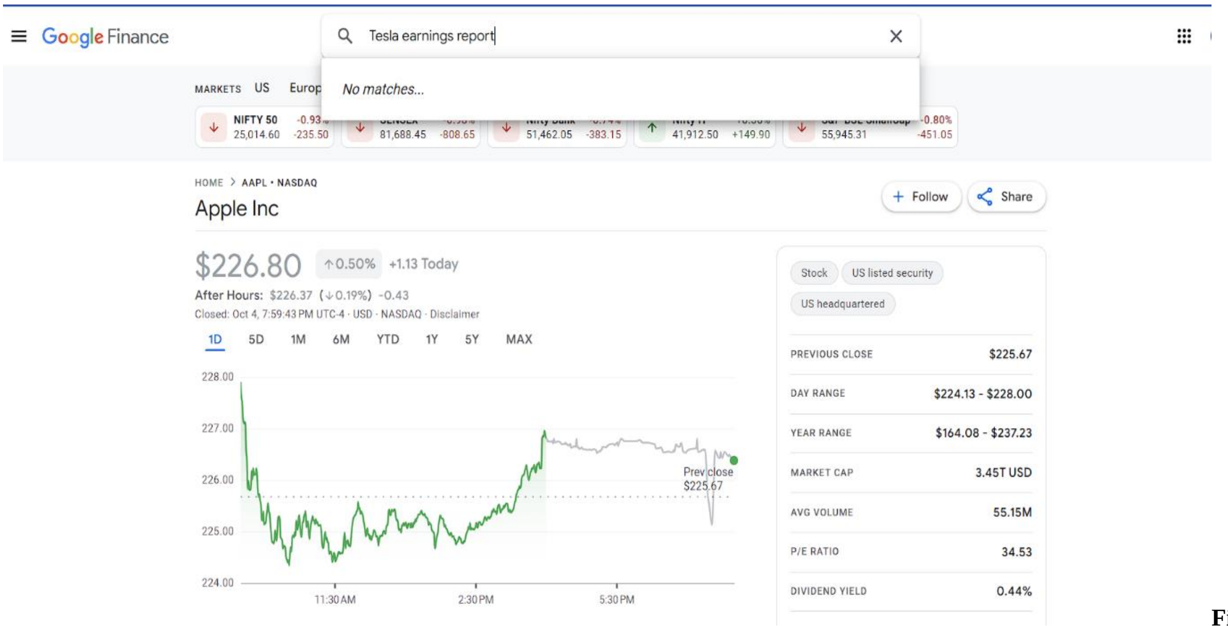

Figure 4. Google Finance search for Deeper research needs "Tesla earning reports".
(Source: https://www.google.com/finance/?hl=en%20&%20https://finance.yahoo.com/.)

5.2 Scenario 2: Deeper Research Needs

User on System 1: Search for "Tesla earnings report." The system will display any recent news articles that have those keywords. You might need to filter through various articles to find the necessary information. Or it may not display an error because of the predefined data retrieval algorithms (see Figure 4).

User on System 2: Search for "Tesla earnings report impact on the stock price." The system analyzes your query and understands you're interested in the potential effect of the earnings report. It displays relevant news articles, analyst opinions, and charts specifically focused on this aspect. It might even summarize key points from the earnings report and its impact.

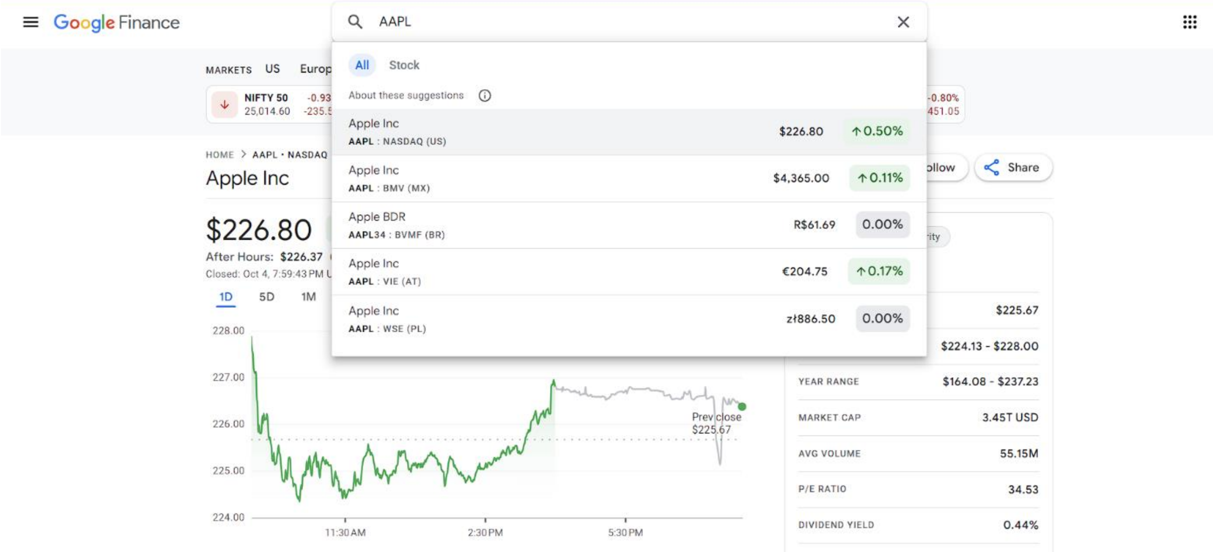

Figure 5. Google Finance search for "AAPL" (Apple INC.).

(Source: https://www.google.com/finance/?hl=en%20&%20https://finance.yahoo.com/.)

5.3 Scenario 3: Personalized Learning and Recommendations

User on system 1: Offers limited user adaptation. The content you see stays the same unless you search for more specific or detailed terms.

User on system 2: The presented model can be learned from the user search pattern over time and filters its search results. For example, if users search frequently for information about dividend-paying stocks, the system will display companies with high dividend yields or upcoming dividend payouts.

6. CONCLUSION

Comprehending the user's intent behind the search is essential for searching for the relevant result on the database and delivering a user-centric and relevant search experience in a constantly evolving financial world. This study demonstrates the limitations of available financial information retrieval systems, which rely on market data feeds and previously defined keywords. This study put forward a new approach that makes user queries more advanced to generate more personalized results and provide the target information goals of the user.

The introduced model has many advantages, and the system will do many things besides match keywords. It can show results according to the user's search and understand the intent behind every query. Ordinary financial information retrieval systems cannot provide a personalized experience, but the proposed approach can give users more nuanced and relevant results. Also, the system can save users efforts and time invested in their economic and financial search by potentially predicting the user's needs and displaying the most appropriate information.

However, it is essential to consider the challenges related to this approach. User data collection and privacy concerns matter most and require proper handling and consideration. Solid algorithms and effective data processing techniques are essential to efficiently handling vast volumes of queries.

Future research direction, searching for advanced user query analysis techniques, embedding feedback mechanisms, and evaluating the effectiveness of the model in real-world user scenarios by considering presented challenges, can take one step ahead towards the refinement of the model, including this future research can contribute to developing a new generational financial information retrieval system that can be more user-centric and personalized.

7. REFERENCES

[1] Odonkor, B., Kaggwa, S., Uwaoma, P. U., Hassan, A. O., & Ajoke Farayola, O. A. (2024). The impact of AI on Accounting Practices: A Review: Exploring how Artificial Intelligence transforms traditional accounting methods and financial reporting. *World Journal of Advanced Research and Reviews*, 21(1), 172–188. https://doi.org/10.30574/wjarr.2024.21.1.2721

[2] Domain Specific Language Models. Lark. (2023). https://www.larksuite.com/en_us/topics/ai-glossary/domain-specific-language-models

[3] Allan, J., Aslam, J., Belkin, N., Buckley, C., Callan, J., Croft, B., Dumais, S., Fuhr, N., Harman, D., Harper, D. J., Hiemstra, D., Hofmann, T., Hovy, E., Kraaij, W., Lafferty, J., Lavrenko, V., Lewis, D., Liddy, L., Manmatha, R., … Zhai, C. (2003). Challenges in information retrieval and language modeling. *ACM SIGIR Forum*, 37(1), 31–47. https://doi.org/10.1145/945546.945549

[4] G. Penha, E. Palumbo, M. Aziz, A. Wang, and H. Bouchard, "Improving Content Retrievability in Search with Controllable Query Generation," in ACM Web Conference 2023 - *Proceedings of the World Wide Web Conference*, WWW 2023, Association for Computing Machinery, Inc, Apr. 2023, pp. 3182–3192. doi: 10.1145/3543507.3583261.

[5] Gidofalvi, G., & Elkan, C. (2001). Using news articles to predict stock price movements. Department of computer science and engineering, university of California, san diego, 17.

[6] Schumaker, R. P., Zhang, Y., Huang, C.-N., & Chen, H. (2012). Evaluating sentiment in financial news articles. *Decision Support Systems*, 53(3), 458–464.

https://doi.org/10.1016/j.dss.2012.03.001

[7] Wang, D., Su, J., & Yu, H. (2020). Feature extraction and analysis of natural language processing for deep learning English language. IEEE Access, 8, 46335-46345.

[8] Cui, L., Xu, Y., Lv, T., & Wei, F. (2021). Document ai: Benchmarks, models, and applications. arXiv preprint arXiv:2111.08609.

[9] Wu, S., Irsoy, O., Lu, S., Dabravolski, V., Dredze, M., Gehrmann, S., ... & Mann, G. (2023). Bloomberggpt: A large language model for finance. arXiv preprint arXiv:2303.17564.