Know About Company

DIPTESH RAVINDRA VARULE[1], SHWETH GOVERDHAN SHETTY[2], RUSHIKESH SANDEEP BORSE[3], PROF. NAMAN V. BURADKAR[4]

[1, 2, 3, 4] *Department of Information Technology, SCTR'S Pune Institute of Computer Technology, Pune*

Abstract- ***The primary method used by the banking industry to disseminate its financial figures is annual financial reports. These intricate and extensive reports require manual processing to extract meaningful information, which causes delays and uncertainty in investment decisions. It is challenging to create an automated system for answering intelligent financial enquiries. A number of studies have been suggested to use information extraction to solve these issues, however they do not address the semantic interoperability of the reports across various organizations. In order to respond to financial queries utilizing ontology-based information extraction, this work presented an automated querying engine. A Financial Knowledge Graph has been proposed to aid in the semantic modelling of financial reporting. Our project seeks to shorten the amount of time users spend learning about the stakes that shareholders and promoters have in a company. Instead of having to manually search through thousands of files, someone can now quickly find out all the details about a corporation. By employing a simple search query, the user will be able to see not just the most recent trends in the organization's various shareholding patterns, but also historical trends, saving a tonne of time. The final objective is to offer a comprehensive online directory (Knowledge Graph) that will, in a graphical and understandable way, respond to any question the user may have regarding the shareholdings of various organisations and people within a company.***

Indexed Terms- ***Ontology-based Information Extraction, WebScrapping, Knowledge Graph, Machine Learning***

I. INTRODUCTION

Investors find it challenging to read and understand the financial consequences described in the quarterly and annual reports published by corporations because the necessary information is widely spread throughout them. The other issue is that practically every corporation creates a large report that is difficult for the stakeholders to read.

Information extraction from these financial disclosures automatically is challenging because there are no clear divisions between the targets' context dependence, the things to be extracted, and things, differences in linguistic style, and statistical techniques limitations. Another issue with extracting information from these financial datasets is that they are typically provided as unstructured texts or in PDF, which necessitates either painstaking manual preprocessing or the use of sophisticated ETL (Extract, transform, load) tools to automatically ingest data.

In this project, we have used knowledge graphs (KG) for a variety of purposes. First off, unlike relational databases, it may not always contain a semantic layer to explain the entity model. Second, the schema is flexible and adaptable, making it simple to change the schema without affecting other graph entities or adding new properties to an existing record.

Finally, graph databases are able to respond to queries that cover via graph traversal over several entities. Those nodes alone which are made available by the inquiry are investigated by the database engine for graphs. given that each record is handled When used alone, it significantly improves query performance and reduces the cost of the query results' resources. An investor will receive help from our knowledge graph by receiving financial stories that can help with decision making. The next step is to create text-based financial narrative and expand it to include visualisations and images.

II. LITERATURE SURVEY

While reading through multiple different papers and websites we found that there were a few common problems that the people were facing like the stock market of the renowned company changes is random, complex, and unstable given that they are affected by many factors, so predicting and monitoring all has become tough. Another problem is that Annual Financial Reports are the core in the Banking Sector to publish its financial statistics. Extracting useful information from these complex and lengthy reports involves manual process to resolve the financial queries, resulting in delays and ambiguity in investment decisions. The papers mentioned various solutions of building complex databases by using various data extraction methods, using natural language processing for extracting valuable information from text documents and various feature extraction methods for making prediction models. Hence the idea of forming a online directory of all the shareholding data was formed and the decision was made to make use of Knowledge Graphs. First off, unlike relational databases, it may not always contain a semantic layer to explain the entity model. Second, the schema is flexible and adaptable, making it simple to change the schema without affecting other graph entities or adding new properties to an existing record. An investor will receive help from our knowledge graph by receiving financial stories that can help with decision-making. The next step is to create text-based financial narrative and expand it to include visualisations and images.

III. PROPOSED METHODOLOGY

A. Data extraction using Selenium

Since Know about Company will be a platform to query and analyze promoter and shareholder data the first step is developing a web scraper which can scrap all the data from the data sources which in our case is the BSE website. The website is updated with the promoter data every quarter end, with all the important shareholding data about every listed company on the Bombay Stock Exchange. This is the first and one of the most important modules as the most important part of the application is to represent accurate shareholding information.

B. Information Extraction

The process of creating knowledge graphs, which addresses the issue of how to automatically extract potential knowledge units from heterogeneous information from several sources, begins with information extraction. Named Entity Recognition refers to the extraction of multiple specified types of entities from unstructured or semi-structured text information. Relation extraction To obtain the association relationship between enterprise entities, semantic information can be obtained through relationship extraction.

C. Knowledge Fusion

Following information extraction, we obtained enterprise entities, relations between entities, attributes and some time information. To improve the logic and level of knowledge, it is necessary to integrate knowledge to improve the quality of the knowledge. Entity linking and knowledge merging are the two main tasks of knowledge fusion.

D. Knowledge Retrieval

The final step after we have the Knowledge Graph is to make sure that we retrieve data properly from it such that the system can answer any type of query the user has regarding the promoter and shareholder details of any company. This data once retrieved should also be properly represented to the user in an easy to understand and graphical manner.

IV. SYSTEM ARCHITECTURE

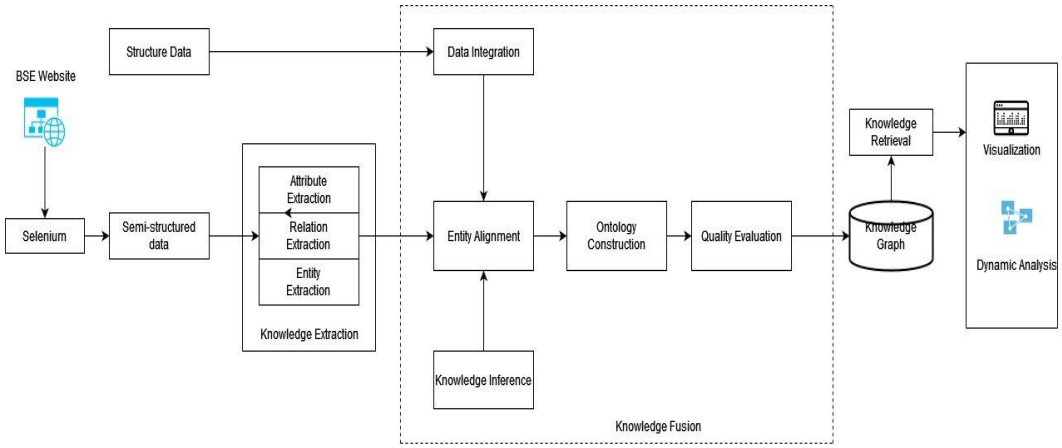

Fig. 1. KAC System Diagram

The Fig.1 shows the complete system architecture of the Know About Company application. This includes all the four proposed methodology that is the data extraction, information extraction, knowledge fusion and Knowledge retrieval.

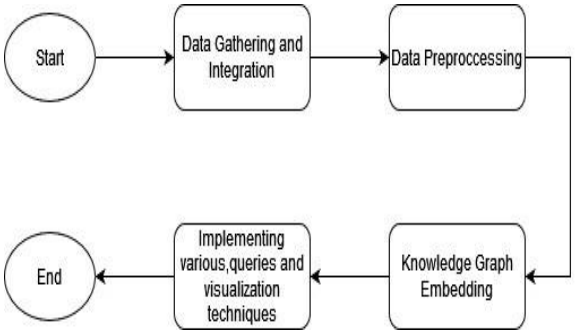

Fig. 2. Data Flow Diagram

The data extraction is done using the selenium framework which automatically queries the Bombay Stock Exchange website for all the previous year quarterly reports and extracts the different shareholding data which are present in various sections of the reports. Once this step is done the Knowledge Extraction begins where the data scraped from the Bombay Stock Exchange website goes from various procedures and data pre-processing procedures to abstract the attributes, the main entities and the relationships between these entities so here the different promoters and shareholders are identified and the relationships between them and the different companies are established.

This is followed by the step of embedding these entities and relationships into the graph database such that they form a proper knowledge graph which contains meaningful data regarding promoters, shareholders and the different listed companies from which we can gain some valuables insights. The final step is retrieval of this knowledge from the Knowledge Graph and visualizing it and representing it to the user according to their queries such that it is easy to understand and learn from.

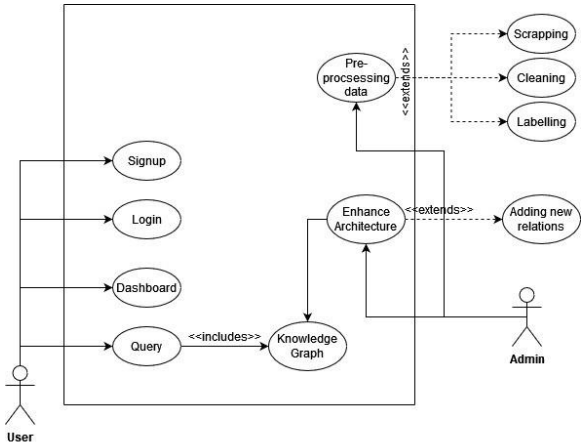

Fig. 3. Use Case Diagram

V. CONCLUSION AND FUTURE SCOPE

A. Conclusion

This study put out a fresh method for gathering information from annual reports of companies in order to populate a financial knowledge graph. We

spoke about the methods used. For information extraction, the engineering of ontologies, the building of a financial knowledge graph, and a question answering system for the graph. Companies are required to publish their annual reports online by financial regulatory bodies in the majority of the world's nations. Despite the fact that the essential dataset is present in this report, our investigation required much effort because the format was in PDF. The report included several sections, therefore automatic extraction took a lot of time and work. Additionally, different companies can use different entity identifiers and formats for the same type of data. The outcome demonstrates that our system, which is based on a financial knowledge graph, successfully answers questions from the general public or investors regarding shareholding data in several companies or for a specific company, efficiently and without any hiccups.

Our goal is to create a web crawler to collect shareholder data from legitimate web pages, a generalisation of the model for businesses other than banks, an automatic entity linking and disambiguation mechanism for knowledge graph (KG) enrichment, the creation of financial stories for various types of stakeholders that can aid decision making, and the exploitation of open information standards.

B. Future Scope

Only companies are taken into account in this study. This study can be applied in a variety of healthcare, travel, petroleum companies, and production organisations, etc. that produce yearly reports however, their clients and end users cannot inquire or gather the desired details quickly.

VI. ACKNOWLEDGMENT

It gives us great pleasure in presenting the preliminary project report on Know About Company. We would like to take this opportunity to thank our guide Prof. Naman V. Buradkar for giving us all the help and guidance we needed. We are really grateful to him for his kind support. We are also thankful to our reviewers Prof. Jagdish Kamble and Prof. Sandeep Warhade for their valuable suggestions. We are grateful to Dr. Archana S. Ghotkar, Head of Department of Information Technology, Pune Institute of Computer Technology for her indispensable support and suggestions.

REFERENCES

[1] B. Zhang et al., "Similarity Analysis of Knowledge Graph-based Company Embedding for Stocks Portfolio," 2021 *IEEE 6th International Conference on Smart Cloud (SmartCloud)*, 2021, pp. 84-89, doi: 10.1109/SmartCloud52277.2021.00022.

[2] Zehra, MS Mohsin, Syed Wasi, Shaukat Jami, Syed Siddiqui, Muhammad Shoaib Syed, Khaliq-Ur-Rahman Raazi. (2021). Financial Knowledge Graph based Financial Report Query System (January 2021). *IEEE Access*. PP. 1-1. 10.1109/ACCESS.2021.3077916.

[3] Liu, Yang Zeng, Qingguo Ordieres-Mere, Joaqu´ ´ın Yang, Huanrui. (2019). Anticipating Stock Market of the Renowned Companies: AKnowledge Graph Approach. Complexity. 2019. *10.1155/2019/9202457*.

[4] Q. Wang, Z. Mao, B. Wang, and L. Guo, "Knowledge graph embedding: A survey of approaches and applications," *IEEE Trans.* Knowl. Data Eng., vol. 29, no. 12, pp. 2724-2743, Dec. 2017, *doi: 10.1109/TKDE.2017.2754499*.

[5] S. Choudhury, K. Agarwal, S. Purohit, B. Zhang, M. Pirrung,W. Smith, and M. Thomas, "NOUS: Construction and querying of dynamic knowledge graphs," in *Proc. IEEE 33rd Int. Conf. Data Eng. (ICDE)*, Apr. 2017, pp. 1563-1565.

[6] Z. Chen, S. Yin, and X. Zhu, "Research and implementation of QA system based on the knowledge graph of chinese classic poetry," in *Proc. IEEE 5th Int. Conf. Cloud Comput. Big Data Analytics (ICCCBDA)*, Chengdu, China, Oct. 2020, pp. 495-499, *doi: 10.1109/ICCCBDA49378.2020.9095587*.

[7] Q. Wang, Z. Mao, B. Wang, and L. Guo, "Knowledge graph embedding: A survey of approaches and applications," *IEEE Transactions on Knowledge and Data*

Engineering, vol. 29, no. 12, pp. 2724–2743, 2017.

[8] H. Paulheim and P. Cimiano, "Knowledge graph refinement: A survey of approaches and evaluation methods," *Journal of Web Semantics: Science, Services and Agents on theWorldWideWeb*, vol. 8, no. 3, pp. 489–508, 201