# Knowledge Graph and Similarity Based Retrieval Method for Query Answering System

**Aaditi Narkhede [1], Prof. M.U.Kulkarni [2], Prof. S.R.Naik [3]**

[1] *Student, aaditirajeshn99@gmail.com, Dept. Of Computer Science and Technology, VJTI, Mumbai, India*
[2] *Professor, mukulkarni@ce.vjti.ac.in, Dept. Of Computer Science and Technology, VJTI, Mumbai, India*
[3] *Professor, sraksha@it.vjti.ac.in, Dept. Of Computer Science and Technology, VJTI, Mumbai, India*

---***---

**Abstract -** *A semantic network that displays the connections between entities is known as a knowledge graph. Data can be visualised to help with information analysis and comprehension through the use of the knowledge graph. Knowledge graphs can help professionals with complex analysis applications and deci- sion support while also erecting barriers in the market. A good construction system can help companies construct knowledge graphs efficiently and quickly. However, Extraction of valuable information from such huge, complex and unstructured data requires a human approach to handle user queries related to data, which causes delays and uncertainty in decision making and strategy planning. In this paper, we propose an approach for au- tomatic knowledge graph construction and automated querying engine to answer user queries using generated knowledge graph. The proposed system will benefit Professionals with faster, easierunderstanding and analysis of complex and huge unstructured data. Experimental results show that our proposed solution is more effective in constructing a generic knowledge graph.*

***Key Words***: **Knowledge graph, Question Answering system, spaCy, Natural Language Processing, Named Entity Recognition.**

## 1. INTRODUCTION

The goal of a knowledge graph is usually to collect, connect, and show information. It offers a high level of inter- pretability. Knowledge Graphs help in establishing purposefulrelationships of organisational knowledge through classifying various content into different categories [1]. By grouping different types of content into distinct categories, knowledge graphs assist in establishing intentional relationships between organisational knowledge. Large volumes of unstructured data are produced daily. This information is presented in reports, research papers, patents, scholastic articles, book chapters, essays, and speeches, among other formats. Identifying key patterns in vast amounts of unstructured data is crucial in to- day's environment. It is challenging to understand and evaluatethe implications provided in an organization's data since the essential information is dispersed across the large volumes of data.

Due to the absence of boundaries between the items that need to be retrieved, the target entities' context dependency, the variability in language patterns, and the limits of statisticalapproaches, automatic information extraction from such vast amounts of data is challenging. The fact that this type of datais frequently available as unstructured texts or in PDF format presents another challenge when trying to extract information from it. As a result, either laborious manual preprocessing is required or sophisticated ETL (Extract, transform, load)systems are used to automatically ingest data. To handle this challenge, required data i.e. textual data will be first extracted from PDFs for our research work using Fine Tuned detectron2based model and pytesseract ocr and will be stored in a text filewhich will be used further for information extraction purposeswhile building Knowledge Graph.

The proposed approach aims to resolve issues of ambigu- ity, abbreviations and semantics of text while constructing a knowledge graph, achieved with the use of spaCy based NER for extracting the entity-pairs and relations from the data. Based on triplets obtained while information extraction, the Knowledge graph is constructed. In the proposed Question- Answering system, for Query analysis a similar spaCy based approach is used for entity-pair and relation extraction from user query. In the Answer Extraction module, the combination of approaches is used, such as information retrieval based on the feature information of relevant entities in sentences and uses trained feature classifiers to sort the candidate answers and obtain the solutions; along with matching query triplets with knowledge graph database using generic linguistics rules designed to obtain the solutions.

### 1.1 Organization Of the Paper

Section 2. of the paper describes the previous and current study being carried out in the field of Knowledge Graphs and Question-Answering Systems. It also states the drawbacksand the problems faced in existing approaches. Section 3.of the Paper explains the proposed methodology and the step by step execution of the same with the help of a few examples. Section 4. of the Paper summarizes and analyses the Results obtained and gives an insight regarding how the proposed system can be deployed for Different domain-data. Section 5. of the Paper concludes the Research regarding the generic approach for Knowledge graph Construction and query answering system. Section 6. of the Paper summarizes the References used/followed for the survey purpose.

## 2. RELATED WORK

Our work is inspired by two threads of research: Key techniques in KG construction and Key techniques in QA system.

### 2.1 Key Techniques in KG Construction

A typical knowledge-graph construction process consists of three main components: information extraction, knowledge fusion, and knowledge graph building.

**Information Extraction :** The goal of information extraction is to find and separate entities in a data source, as well as their qualities and relationships with other entities with the means of entity recognition and relation extraction.

Entity extraction, also known as named entity recognition (NER), refers to the process of identifying accurately named entities from data, especially text data [7]. There are certain paradigms that have been developed for NER activities such as Rule based, Machine learning based and Deep learning based approaches. The underlying concept behind most common NER approaches was to create a small set of rules by hand and then look for strings in the text that fit these criteria. Usage of Rule based NER can not handle multiple unforeseen patterns and thus does not possess strong interpretability of semantics while information extraction [2]. Traditional NER approaches, such as rule-based and template-based, are expensive and rely too much on procedures like rule development and feature engineering. The machine learning based approach, tags the named entities to the words even when the words are not listed in the dictionary and the context is not described in the rule set. K. Khadilkar, S. Kulkarni and S. Venkatraman [1], proposed a method for summarizing multilingual vocal as well as written paragraphs and speeches, using semantic Knowledge Graphs. For the selection of relation extraction API, experiments were performed on NLTK and Stanford CoreNLP; out of which Stanford CoreNLP resulted with highest accuracy of 92.23% which uses CRF algorithm internally having ability to consider context of previous text in detecting entities and relation; but fails to consider context level dependency in relation extraction. Different from traditional machine learning, deep learning can automatically extract high-level abstract features from a large amount of data to perform model training [4]. In particular, word vector representation has provided a powerful driving force for the typical serialized labelling problems of NER. Z. Dai, X. Wang has discussed various neural network models [3], where the BERT-BiLSTM-CRF model achieved higher F1 score compared to other models. BERT, captures a general language representation from large-scale corpora, but lacks domain-specific knowledge. Two approaches such as SpaCy and BERT are discussed in [6] for NER task for Tourism dataset where experimental results shows that SpaCy NER outperforms BERT NER as achieved accuracies are 95% and 70% respectively. Performance of BERT is lowered as it is unable to tokenize special multi word names which are proper nouns very well. Syntactic Dependency Parsing is also offered by SpaCy's dependency parser [6] for relation extraction.

Relation Extraction is a major field in NLP. Using Relation Extraction, data can be transformed into a 3-tuple format of [entity, relationship, entity]. The patterns are generated through text analysis and represent the unique language constructions which are used to describe a particular Entity/Relation in [2]. These patterns are then matched with processed text to discover and extract required pieces of information; extracted relationship names between two entities using object property of an ontological concept. Recognizing the relationships using the spaCy model may be done in two ways. The first method is to utilise spaCy's dependency parser. The second method is to create new tags based on relation keywords. [6] showed comparison between spaCy's dependency parser and BERT with BIO tagging where spaCy achieved higher accuracy of 95% and also it was able to consider context level dependency in relation extraction.

**Knowledge Fusion :** In Knowledge fusion, last step of knowledge graph construction, Entity Disambiguation and entity linking is carried out to preserve the semantic information of text and handles uneven knowledge expression. During the Entity Disambiguation job, ambiguous entity mentions are connected to their referent entities in the Knowledge graph, which is accomplished by employing a clustering technique in [9]. A method based on Word2Vec cosine similarity calculation is proposed in [8] to complete entity linking for entities obtained from question sentences. Word2Vec is used in [8] to convert recognition entity and candidate entity into corresponding word vectors and implemented cosine similarity method to calculate similarity value between recognition entity and candidate entity. For the Question-answering system in [11], relevance scoring based on QA context is used to link and find relevant entities from KG. To link the primary entities with their preposition occurrences in the text or their multiple reference entities, coreference resolution is used [1]. For topmost winning models described in [7], Coreference resolution is achieved using NeuralCoref offered by spaCy. Rather than adopting comprehensive coreference resolution, [12] tackle the problem by recognising the pronouns in the input content and replacing them with the corresponding subject or object. further, accuracy of coreference resolution is improved by analysing the gender of an entity. For the coreference resolution task, it has been observed that NeuralCoref by spaCy does the job in an efficient way.

### 2.2 Key Techniques in QA System

Question-answering research based on Knowledge graphs is an important area of study. The basis of the knowledge graph's question answering mechanism is question analysis. The accuracy with which the semantic information in the question sentence is mined impacts the actual effect of the question answering system. Research based on financial knowledge graph using ontology proposed Question Answering system [2] to sequentially perform linguistic analysis of query, do named entity extraction, entity / graph search, fusion and ranking of possible answers.SPARQL is used for querying.

To solve the problem of inadequate question's semantic information mining, [8] proposed BERT based KG QA system which mainly includes Entity recognition and Relation Recognition for question analysis where BERT-BiLSTM-CRF model is used and calculates the similarity between the entity obtained in the entity recognition and candidate entity, and takes the entity with highest similarity as the standard entity contained in the natural language question.

The mainstream implementation methods of a knowledge based question answering system can be divided into categories such as Semantic Parsing which involves some linguistics and traditional NLP methods, requires many manual design rules, and has high accuracy but lacks generalization ability; and Information Extraction: This kind of approach extracts the feature information of relevant entities in sentences or knowledge bases and uses trained feature classifiers to sort the candidate answers and obtain the solutions. It is closely related to the traditional NLP method and feature engineering, with strong generalization ability but relatively weak accuracy. The combination of these techniques is used in [9]. The key entities and relationships are derived from the user's natural language questions, which are categorised and evaluated. Finally, Cypher language is built to query the knowledge graph via intention prediction. It used question and answer matching technology to calculate the semantic similarity score between the original problem and each candidate solution, choose the best answer based on the highest score, and then return the user's response directly. Experimental results of [9] shows that Question-Answer pairing achieves higher accuracy than simple KG retrieval.

[12] introduces a graph-based QA system for reading comprehension tests that pick out the sentence in the passage that best answers a given question by extracting the relations. The proposed system consists of three main modules –Document Processing, Query Processing and Answer Extraction. In the Answer Extraction module, comparison between generated graph from Document processing and query triplets obtained from Query processing is done to determine a set of matching sub-graphs. A morphological analysis which includes a tense variant check for the verb is carried out if no match occurs in the comparison step which improves the accuracy of the model. model achieve accuracy of 79.67%

### 2.3 GAP Analysis

The challenges in knowledge graph construction are :

1) information loss

2) information redundancy

3) information overlapping.

4) Knowledge noise

Information loss occurs when information extraction is poor, resulting in an incomplete output graph. Information redundancy refers to the repetition of the same entity with different abbreviations or prepositions, as well as extra concepts and relations that do not exist in the input text but do exist in the background knowledge. The information overlapping challenge refers to whether a knowledge graph can encode the changing of an attribute. Too much knowledge incorporation from LMs may divert the sentence from its correct meaning, which is called knowledge noise (KN) issue.

Several approaches on which these systems are based, the best known of which are based on Information Extraction using Named Entity Recognition approaches for knowledge graph construction and Traditional query processing and pattern matching approaches for Question-answering systems. The hybrid approach is an alternative trying to merge the advantages of these methods to fill the weak points.

## 3. METHODOLOGY

We have divided the research work broadly in certain phases. All the phases are discussed separately in this paper.

### 3.1 Data Acquisition and Processing

The Knowledge Graph and Question-Answering System constructed in this paper is based on the Annual Reports of certain Companies such as Apple, Facebook, ACC and RIL, etc which discusses various financial statements, listed company executive information, news, announcements and research reports. In this research we are considering only textual data for KG Construction. These Reports are generally in unstructured format, PDF formats and also consist of various graphics, charts, images, tables and text data which are in a highly unstructured format; making it difficult for extraction of data from PDF reports.

To handle this, we extract the required textual data for KG construction into a single text file for all companies.

For the Data extraction, we use Layoutparser which performs Document Image Analysis with the help of state-of-the-art detectron2 deep learning model which enables identify and extract complicated document structures.

A custom dataset is being created and annotated using labelImg tool by giving annotations for portions of document

to be extracted and fine tuned the pre-trained faster rcnn detectron2 model on our custom dataset and annotations. Finally, with the use of Fine Tuned model, Inferences are made and extracted outputs are saved in a single text file which will act as an input data for the construction of KG and QA System.

### 3.2 System Architecture

The proposed System architecture, which presents how the user queries are processed and how the system will generate the results, is shown in Fig. 1.

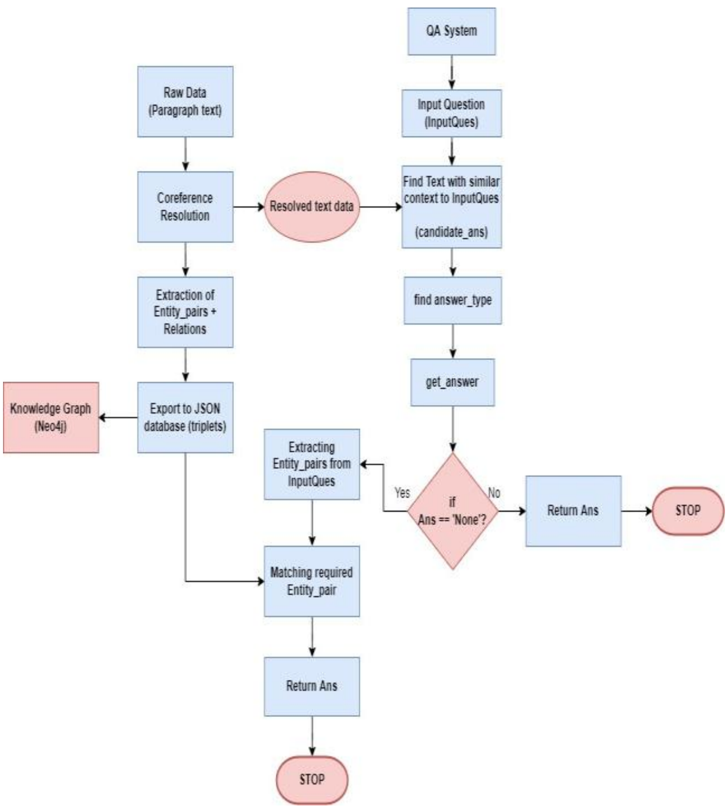


Fig. 1. Proposed System Architecture

### 3.3 Coreference Resolution

**Input Sentence :** "' *The ACC company follows a robust process of communicating with its stakeholders and investors. For this purpose, it provides multiple channels of communications through dissemination of information on the on-line portal.* '"

**Coref clusters:** [The ACC company: [The ACC company, its,it]]

**Resolved text:** *The ACC company follows a robust process of communicating with The ACC company stakeholders and investors. For this purpose, The ACC company provides multiple channels of communications through dissemination of information on the on-line portal.*

Fig. 2. NeuralCoref identifies the coreferences for an entity *The ACC Company* in the above example and resolves them by replacing with an entity name.

To connect the primary entities with the prepositional occurrences of those entities in the text, coreference resolution is used. With this process pronouns get replaced with proper nouns thereby generating more interpretable graphs. Additionally, it is employed when a knowledge base object is connected to several entity references. For instance, since "President Modi" and "Narendra Modi" relate to the same person, they should be combined before being linked to an entity in the knowledge base. Coreference Resolution is also a heavy NLP task. We have used NeuralCoref; a pipeline

extension for spaCy 2.1+ which annotates and resolves coreference clusters using a neural network. It is integrated in spaCy's NLP pipeline. NeuralCoref has been proved to be both more efficient and accurate for coreference resolution tasks. Example of working of NeuralCoref pipeline is shown in Fig. 2.

### 3.4 Information Extraction

The second phase in creating a knowledge graph is information extraction. The main challenge is identifying candidate knowledge units by autonomously extracting data from data sources. Information extraction is challenging since natural language processing (NLP) technology is typically required when working with semi-structured or unstructured data. Entity and Relation Extraction are among the important technologies.

The purpose of entity extraction, sometimes referred to as named entity recognition (NER), is to create "nodes" in knowledge graphs.An essential component of information extraction, entity extraction has a significant impact on the effectiveness and calibre of later knowledge acquisition. Entity class—which includes names of people, places, and insti- tutions—time class—which includes dates and times—and number class—which includes terms like money and percent- age—are the three basic classes. It is possible to expand these classes to accommodate various application areas. In this step,we extract the pairs of entities which are related to each other from each sentence of the data. We employed the NLP library SpaCy, in order to categorise words with their appropriate part-of-speech (POS) tags and chunked noun and verb phrases extracted in accordance with specified rules, which are able to consider all dependent noun and verbs in a sentence with the help of dependency parser offered by spaCy. A verb chunk is defined as the verb and any accompanying adpositions or particles, whereas a noun chunk is defined as the words characterising the noun.

After obtaining the entities i.e. nodes in a graph using entity-pair extraction, the process then moves on to relation extraction for edge construction. Extracting the relationship i.e. edge between entities is required to gain semantic information. In order to map pairings of entities, we first extracted relation terms from sentences, such as verbs, prepositions, and postpositions. We then combined each relation phrase with its source and target entities to generate triplets. Extracted triplets by using SpaCy and a set of linguistic rules based on subjects, objects, predicates, and prepositions so that it can work on domain independent data. Along with normal triplet structure i.e. <ent1, relation, ent2>; we have additionally extracted auxiliary relation, time and place such as <source, relation, aux rel, target, time , place> making it better for further QA System.

### 3.5 Knowledge Graph

We build knowledge graphs using a bottom-up strategy,

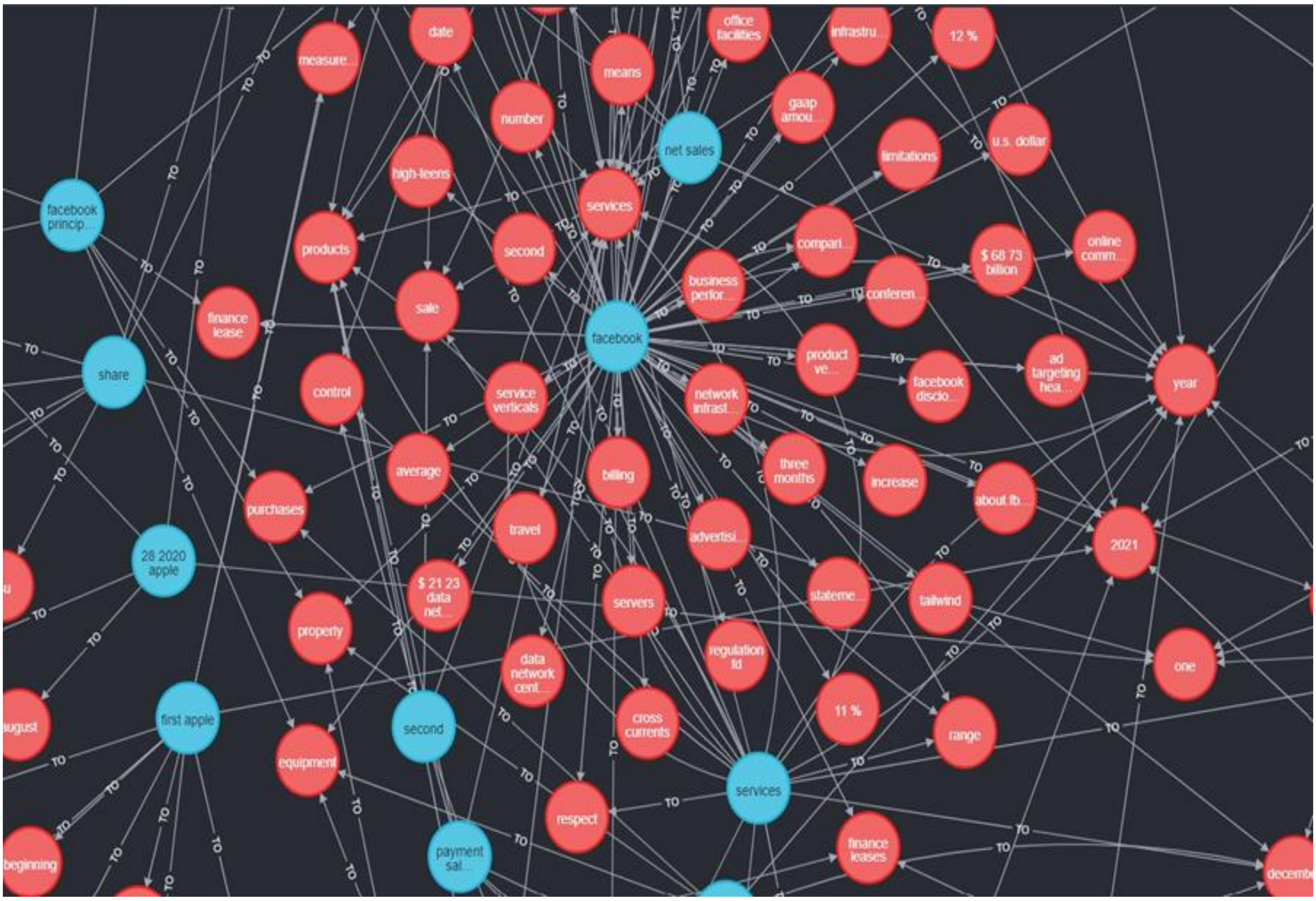


Fig. 3. Visualization of KG for Annual Reports ( with zoom view)

where entities and relationships are first pulled from the data and the upper model layer is generated via data driving. After entities are recorded in triple form, the knowledge graph can then be visualized using the Neo4J graph database. First, we encode the triplets that were recorded into a CSV file using UTF-8, and then we used the Cypher import command LOAD CSV to put the triplets' data in the CSV file in accordance with the triplet "ENTITY —[r:RELATIONSHIP]-> ENTITY" in the graph database. For the Visualization of a KG for Annual reports following Cypher query is processed :

"' LOAD CSV WITH headers from "file:///Triplets Database.csv" as row with row where row.source is not null merge (n:Source {id:row.source}) merge (m:Target {id:row.target}) merge (n)-[:TO{rel:row.relation}]- (m) return * "'

As the input triplet data is large, the generated KG is huge. Here, Neo4j comes as the best solution as it allows users to traverse and navigate through the entire KG and visualize and interpret various properties and attributes of nodes and edges by hovering over them. Visualization results for KG for Annual reports are shown in Fig. 3.

We can Visualize triplets only for the specified relation, for example to visualize triplets only for relation 'excelled' using following query :

"'match (n)-[:TO{rel:'excelled'}]->(m) return *"'

Results of above query can be seen in Fig. 4. Also by hovering over the edges and nodes we can display and interpret their attributes and properties.

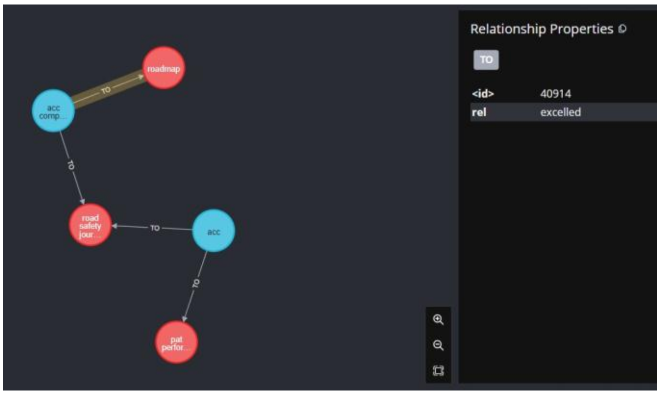


Fig. 4. Visualization of Relation by hovering over edges.

### 3.6 Question-Answering System

The question-answering system suggested in this paper is built using a combination of two approaches: semantic parsing, which involves some linguistics and traditional NLP methods, requires rules; and another is extracting the feature information of pertinent entities in the question and finding matching triplets for the query from knowledge databases to obtain the solutions. The primary entities and relationships are determined by analysing the user's natural language queries. Unlike other generalized QA systems which usually works only for processing Factoid-type Questions ('Wh' type objective questions), The proposed QA System can work for Factoid type as well as Descriptive type of Questions such as questions that start with the keywords of "why" and "how".

1) Query Analysis and Processing : In relation to the

context of the query, the similarity computation is performed to identify comparable sentences from data. In this paper, the Cosine similarity discrimination method is adopted. The recognition entity and candidate entity must be transformed into corresponding word vectors in order to assess how similar they are to one another. Cosine similarity is calculated by the angle between two vectors, given as follows :-

$$cos(\theta) = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}}$$

Two words' vector directions are nearly identical if they are synonyms. Their vector directions are essentially the opposite if they are antonyms. As a result, the criteria are the same if the two words' semantics are more similar and their vector directions frequently coincide.

Term Frequency - Inverse Document Frequency is referred to as TF-IDF. A numerical statistic called the TF-IDF rates the significance of each word in a document. To count the word occurrence in each document, we use TfidfVectorizer functions that are provided by Scikit-Learn library. Term Frequency: Number of times a word appears in a text document. Its formula is as follows :-

$$TF_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}}$$

Inverse Document Frequency: Measure the word is a rare word or common word in a document.

$$IDF_{i,j} = log\left(\frac{D}{1+j}\right)$$

Finally, multiplying them can obtain TF-IDF values,

$$TF \bullet IDF_{i,j} = TF_{i,j} * IDF_{i,j}$$

Using the cosine similarity, the indexes of text with similar context as that of question are returned which gives the sentences which are supposed to be candidate answers. Once the candidate answers are obtained, the Query processing is carried out.

In the Query Processing part, Input Query is processed to find out the expected answer type according to the question using POS tags and NER tags of the question sentence. With the help of various linguistics based design rules, the task of determining required entity type is carried out such as in short way, following describes the design rules for finding answer type for required solution :

"WHO" : "PERSON" or "ORG"

"WHEN" : "TIME"

"WHERE" : "GPE"

"HOW MUCH" : "MONEY"

"HOW MANY" : "QUANTITY"

"HOW" : "DESCRIPTION"

"WHAT" : "DESCRIPTION"

"WHY" : "DESCRIPTION"

2) Answer Matching: Once the answer type is estimated, for Answer Extraction, the corresponding equivalent entities are searched from candidate answers and returned as the final solution to the query. In case, if the system cannot identify the correct candidate answer and hence fails to extract the required entities, then System will move forward to knowledge database retrieval. In the Knowledge database retrieval method, first the triplets are extracted from a question which defines what information the user is demanding. Then the answer is extracted by triplet matching where the question triplet is traversed and matched over the triplets database. For Descriptive Kind of questions, the candidate answer found with usage of cosine similarity feature is returned as the solution by system.

## 4. RESULTS AND ANALYSIS

We implemented the Knowledge Graph and Question answering System for the Annual Financial Reports of various organizations, such as Facebook, Apple, RIL, and ACC,etc. We have constructed a single Knowledge graph for textual data comprising data from all company's Reports. In the Knowledge graph Construction Process, the proposed system is able to identify and extract a total 77,575 triplets comprising entity-pairs and relations. With the use of spaCy's Dependency Parser, it achieved the extraction of all possible entity-pairs and relations between them in a better way for the complex grammar structured data of Annual reports. Fig. 3. Shows the Knowledge Graph generated for Annual Financial reports of Companies.

In the actual use of the QA system, the combination of information retrieval from document and Knowledge graph based Question-Answer triplet matching is used. With the use of these combined approach, Proposed QA achieves the functionality to work for both Factoid as well as Descriptive type of questions. User queries' results are relatively accurate due to high quality data and strong Knowledge fusion (Coreference resolution) steps. Currently, accuracy is used as the basis for the performance test of the question answering system. The accuracy increases with the number of correct questions the question-answering system answers. To handle QA system evaluation's subjectivity, test data is collected from users through Google Forms where users were asked to provide question and answer for given paragraph data. For the Annual Financial reports, the proposed system achieved the accuracy of 81.34% by answering 327 questions correctly out of 402 questions.

For testing the Generalization capability of the proposed system, performance of the system is tested on other-domain data such as Ayurveda and Wildlife Management. PDF data for Ayurveda and Wildlife Management is collected from https://www.ayurveda.com/pdf/intro ayurveda.pdf and https://old.mgkvp.ac.in/Uploads/Lectures/49/859.pdf respectively. It has been observed that the system is able to identify and extract all possible entity-pair and relations. For Empirical verification, Question-Answer dataset is collected from users for Ayurveda and Wildlife management domain through the Google Forms. Following table summarizes the results :

**Table -1:** Results

| Domain | Correctly Answered Questions | Total Questions | Accuracy |
|---|---|---|---|
| Ayurveda | 128 | 150 | 85.33% |
| Wildlife Management | 121 | 150 | 80.66% |
| Financial Annual Reports | 327 | 402 | 81.34% |

Accuracy of 85.33% and 80.66% is achieved for other domains, Ayurveda and Wildlife Management data respectively.

For the comparative evaluation, we compared our system with existing system [12], which uses a graph-based approach for answer retrieval for QA system handling reading comprehension tests with accuracy 79.67%. Existing system performs the Document processing and Query Processing parallely. Graph is generated for triplets obtained from the document, extracts the triplets from user questions using spaCy and at last answer is extracted by subgraph matching. So, the existing system fails if the exact matching subgraph is not found. Also, it works for Factoid-type questions only and can't answer descriptive questions. To address these issues, the proposed system uses cosine similarity feature to find the candidate answers that match the context of questions along with the answer retrieval based on knowledge database triplet matching and addresses descriptive questions as well.

However, the proposed system fails to answer the analytical and confirmation (yes/no) type of questions. Fig. 5 shows the summary of comparative analysis for graph-based(existing system) and Combination of Knowledge graph QA along with retrieval based on the feature information(proposed system).

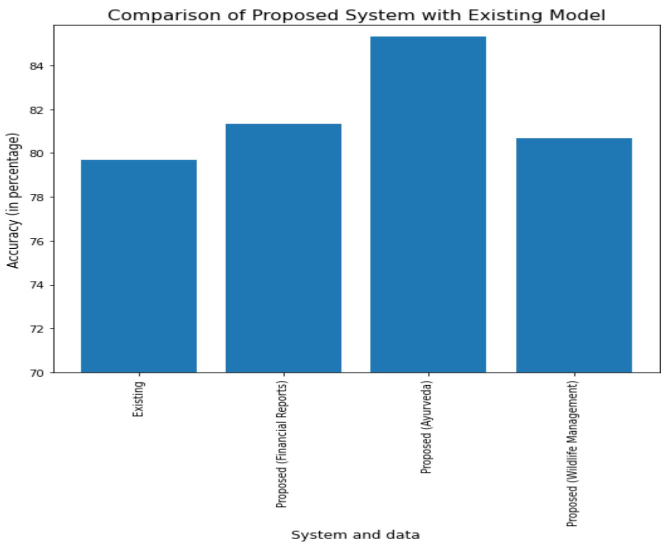


Fig. 5. Comparison graph in terms of Accuracy

## 5. CONCLUSIONS

This system establishes a generalized approach for constructing Knowledge Graph and an automated querying engine to answer user queries for faster and better understanding of huge data. The system resolves the issues of ambiguity and semantics of text by Named Entity Recognition using spaCy which has been observed to be successful for extracting entities from complex structured text. Experimental results show that the proposed approach works efficiently for the different domain data. The combined approach, Knowledge graph QA along with retrieval based on the feature information used in the proposed QA system widens system scope, as it allows the system to process and answer descriptive questions. However, the system generates only one sentence answers for the descriptive queries. Also, the system is unable to answer confirmation and analytical questions correctly. In the future work, more advancement can be done to the system to provision processing of various types of questions and inference based answer retrieval.

## REFERENCES

[1] K. Khadilkar, S. Kulkarni and S. Venkatraman, "A Knowledge Graph Based Approach for Automatic Speech and Essay Summarization," 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), 2019, pp. 1-6, doi: 10.1109/I2CT45611.2019.9033908.

[2] S. Zehra, S. F. M. Mohsin, S. Wasi, S. I. Jami, M. S. Siddiqui and M. K. -U. -R. R. Syed, "Financial Knowledge Graph Based Financial Report Query System," in IEEE Access, vol. 9, pp. 69766-69782, 2021, doi: 10.1109/ACCESS.2021.3077916.

[3] Z. Dai, X. Wang, P. Ni, Y. Li, G. Li and X. Bai, "Named Entity Recognition Using BERT BiLSTM CRF for Chinese Electronic Health Records," 2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), 2019, pp. 1- 5, doi: 10.1109/CISP-BMEI48845.2019.8965823

[4] R. Miao, X. Zhang, H. Yan and C. Chen, "A Dynamic Financial Knowledge Graph Based on Reinforcement Learning and Transfer Learning," 2019 IEEE International

[5] Liu, W., Zhou, P., Zhao, Z., Wang, Z., Ju, Q., Deng, H., Wang, P. (2019, September 17). K-Bert: Enabling language representation with knowledge graph. arXiv.org.

[6] C. Chantrapornchai and A. Tunsakul, "Information Extraction based on Named Entity for Tourism Corpus," 2019 16th International Joint Conference on Computer Science and Software Engineering (JCSSE), 2019, pp. 187-192, doi: 10.1109/JCSSE.2019.8864166.

[7] X. Wu, J. Wu, X. Fu, J. Li, P. Zhou and X. Jiang, "Automatic Knowledge Graph Construction: A Report on the 2019 ICDM/ICBK Contest," 2019 IEEE International Conference on Data Mining (ICDM), 2019, pp. 1540-1545, doi: 10.1109/ICDM.2019.00204.

[8] H. Jiang, B. Yang, L. Jin and H. Wang, "A BERT-Bi-LSTM-Based Knowledge Graph Question Answering Method," 2021 International Conference on Communications, Information System and Computer Engineering (CISCE), 2021, pp. 308-312, doi: 10.1109/CISCE52179.2021.9445907

[9] Z. Jiang, C. Chi and Y. Zhan, ”Research on Medical Question Answering System Based on Knowledge Graph,” in IEEE Access, vol. 9, pp. 21094-21101, 2021, doi: 10.1109/ACCESS.2021.3055371.

[10] Z. Chen, S. Yin and X. Zhu, ”Research and Implementation of QA System Based on the Knowledge Graph of Chinese Classic Poetry,” 2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA), 2020, pp. 495-499, doi: 10.1109/ICCCBDA49378.2020.9095587.

[11] Yasunaga, M. (2021, April 13). QA-GNN: Reasoning with Language Models and Knowledge Graphs for. . . ArXiv.Org. https://arxiv.org/abs/2104.06378

[12] G. Veena, S. Athulya, S. Shaji and D. Gupta, ”A graphbased relation extraction method for question answering system,” 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2017, pp. 944-949, doi: 10.1109/ICACCI.2017.8125963.

[13] Ostendorff, M. (2019, September 18). Enriching BERT with Knowledge Graph Embeddings for Document Classification. ArXiv.Org. https://arxiv.org/abs/1909.08402

[14] K. Guo, T. Jiang and H. Zhang, ”Knowledge Graph Enhanced Event Extraction in Financial Documents,” 2020 IEEE International Conference on Big Data (Big Data), 2020, pp. 1322-1329, doi: 10.1109/Big-Data50022.2020.9378471

[15] Hsu, C. (2019, December 3). Knowledge-Enriched Visual Storytelling. ArXiv.Org. https://arxiv.org/abs/1912.01496

[16] X. Yang et al., ”A General Solution and Practice for Automatically Constructing Domain Knowledge Graph,” 2020 IEEE 6th International Conference on Computer and Communications (ICCC), 2020, pp. 1675-1681, doi: 10.1109/ICCC51575.2020.9344946.