Review and Analysis of Different Approaches to Semantic Level Question Answering and Information Retrieval

Liji S K[1], Muhamed Ilyas P[2]

Sullamussalam Science College, Malappuram, Kerala, India.
liji.s.k[at]gmail.com

Abstract: ***Question Answering and Information Retrieval plays an important application of Natural Language Processing and Data Mining. It aims to retrieve relevant documents for natural language queries. Here we performed a survey on different models for Information Retrieval and Question Answering. And also performed the comparison and analysis of various models. From the literature, we identified that information retrieval systems use the methods from Data Analytics, Natural Language Processing, Machine Learning, and Neural Network etc. Also, we noticed there are many works are done in the English language, but a few works are done in native languages such as Malayalam, Kannada, and Tamil etc.***

Keywords: Natural Language Processing; Information Retrieval; Question Answering; Machine Learning

1. Introduction

The interaction between Computers and humans are achieved by Natural Language Processing techniques and question answering systems. Question answering is an inter-disciplinary field which uses the techniques from Natural Language Processing, Data Mining and Information Retrieval. NLP researchers gather knowledge about how human being understand and process native languages. Nowadays Question Answering and Information Retrieval is almost demanding and growing research field.

Nowadays a huge amount of unstructured data are scattered across the web and it is growing at an exponential rate also very large numbers of people engaged in information retrieval simultaneously. So Information Retrieval from these large volumes of unstructured data using natural languages become a more crucial and challenging task. The relevant Information Retrieval from such a large amount of unstructured data needs knowledge about the semantic information or contextual information.

*I*n this work, we performed a detailed study and analysis of different Information Retrieval systems exists till now and we identified, how the researchers meet the problem, different approaches and methodologies performed, and finally analyzed the results and observed how the approaches fit in different domains. This paper organized as follows. The background of Information Retrieval and Query Processing is explained in section 2, section 3 discusses various methodologies and architecture of Information Retrieval models, section 4 presented the comparative study and analysis of different models, section 5 describes the conclusion and the last section points out the direction for future research work.

2. Information Retrieval

Question Answering and Information Retrieval are used to automatically retrieve relevant answers for users' queries in natural languages. Users can query the system using their own native languages. The system will process the queries and match with the documents and retrieve the relevant results. NLP question answering is the most reliable method for human-computer interaction. Using NLP techniques the native languages like Malayalam, Kannada, Tamil will be analyzed and processed.

Question Answering system accepts the natural language queries from users. Process the queries and convert them into more meaning full or structured forms. Then analyze and classify them. Then match the queries with documents already available and rank the documents using any similarity measures and retrieve the corresponding results [2] [3] [4]. The architecture of Question answering system is shown in Figure1.

The two important aspects of Question Answering are Query Processing and Document Processing. The document processing is a more challenging step because the answer document must be more relevant and meaningful answers for the queries.

3. Architecture and Methodologies of Related Works

The literature reviews related to this survey mainly focuses on the following perspective, Semantic level Information Retrieval and Query Processing, In- formation Retrieval and Query processing in Natural languages and different approaches to Semantic level Information Retrieval. First, we taken all the research papers exists in this field since 2015, it contained about more than

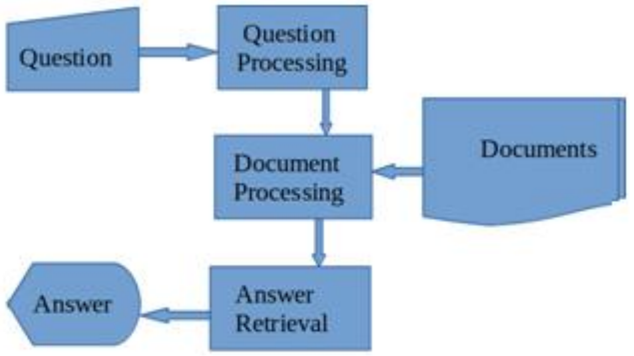

Figure 1: Architecture of Question Answering System

80 papers. After the content filtering reviewed about 26 papers till 2020. The detailed architecture and methodologies of different works are described as follows.

In a work Nadia Soudani et al[5] described an Arabic semantic IR, using a text mining approach. They proposed a generic semantic search approach on Semantic Spaces. They make a comparative experimental study of NLP tools for Arabic and use of linguistic resources thereby the effect of them on the semantic search performance and the importance of the linguistic choices in alienating semantic search engines results. A module for Query Reformulation is integrated with the System based on a knowledge-based approach for Arabic Semantic Disambiguation by use of a dictionary. The process of Word Sence Disambiguation is done based on a Sense Recognition algorithm. Different Semantic Information Retrieval approaches are experimented relying on Semantic Spaces. Tests were made with the use of different Morphological Analyzers and different linguistic resources. The Mean Average Precision for thesystemvariesfrom0.97to7.52.

In another work Shengxian Wan et al [6] proposed new deep neural network architecture for semantic matching with multiple positional sentence representations named MV-LSTM. They use a bidirectional long short term memory Bi-LSTM. Then model the interaction between the representations, using three operations-Cosine, bi-linear and tens or layer. The nusek-max poling strategy for selecting to pk strongest interactions and produce the result by MLP. Learn the model by Back propagation and Stochastic Gradient Descent. They demonstrate the experiment on semantic matching for QA and sentence completion. The analysis shows that the MV-LSTM achieve 11.4% result than the baseline method.

*I*n a work, Saravana kumar Kandasamy et al [7] proposed method to improve open domain question answering. There are two components query processing and document processing. Query processing uses POS tagging, Named Entity Recognition, Parsing, Keyword extraction, Finding synsets, and Similarity measurement to create alternate queries. Document processing use URL weight calculation and Latent Semantic Analysis to correct answer retrieval. The precession of the system is 0.77 and Mean Reciprocal Rank (MRR) is 0.79.

Piyush Arora et al [8] use a query expansion (QE) methods in information retrieval on WebAP dataset. The different approaches they used are Pseudo Relevance Feedback (PRF), using Robertson term selection and Word Embeddings (WE) of query words to address the query document term mismatch issue. The embedding of each word is performed by using a feed-forward neural network by predicting a word by its context. The Normalized Discounted Cumulative Gain (NDCG) of the system is 0.16 and Mean Reciprocal Rank (MRR) is0.36.

Nouha Othman et al [9] discussed a Community Question Answering (CQA) system. They used a word embedding based method to bridge the lexical gap between the questions. Model the semantic information of words in language vector space by using Word 2 Vec model. The questions are then ranked by using cosine similarity. The previous question with high similarity score with the new queried question will be returned and the find the corresponding answer. The Mean Average Precision (MAP) ranges from 0.39 to 0.45 on different models.

In a work, Shenghui Wang and Rob Koopman [10] compared word embed- dings Word2Vec and GloVe with their own Ariadne approach. They used a neural network-based document embedding method, Doc2Vec with Ariadne approach in the context of Information Retrieval on Medline dataset. The average recall of the Doc2Vec and Ariadne methods is 93.3% and 86.3 % respectively. However, they have shown that Ariadne performs equally well as Doc2Vec in a specific Information Retrieval. If the application is to provide contextual information of a word, Ariadne might be a better choice.

Prathyusha Kanakam et al [11] proposed an algorithm to querying the se- mantic web. It uses SPARQL querying language as well as Linked Open Data Quality Assessment (LODQA) for semantic search that converts natural language user's queries to machine-understandable format. The Web Ontology Language (WOL) is used to describe relationships among classes and classifications. Then by using SPARQL to retrieve most accurate results from these ontologies. In this work, the entire approach follows High-Performance Linguistics (HPL) algorithmic process for the proposed system.

In a work, Reshma PK et al[12] proposed a semantic Information Retrieval model for University domain using ontology by the help of Protege. Ontology is used to compare conceptual information across two knowledge bases on the web, it formally describes a list of terms which represent important concepts, such as classes of objects and the relationships between them to represent an area of knowledge. Ontology Web Language (OWL) is used to build ontologies. The different steps for building Ontologies are ontology capture, ontology coding and integration with existing Ontologies. The different tasks are defined classes and class hierarchy, define object properties and then define instance of ontologies, finally querying with DL Query. The precision and recall parameters of the system are evaluated as 87% and 56% respectively.

Pratibha Bajpai et al [13] discussed the development of English to Hindi Cross-Language Information Retrieval (CLIR) system. They experimented the system with Google and Bing search results documents. They used a two-level word sensed is ambiguation model to perform disambiguation of Hindi words to the English language. To optimize the translation and disambiguation model by adding a new valuable component analyzer in the basic CLIR architecture. The MAP of Google and Bing queries are 0.45and 0.35 respectively.

D Thenmozhi et al[14] developed a Tamil- English Cross-Language Information Retrieval (CLIR) system in the agriculture domain, using Ontology and Word Sense Disambiguation. The MAP of the system is 95.36 percent. Sumit Kumar Mishra, V.K. Singh [10] also build a semantic Information Retrieval system for legal cases using Ontology merger and extended GAIA methodology, which contains information about legal cases. This model provides reasoning capability too.

Piyush Mital et al [15] proposed a graph- based question answering system on Wikipedia documents. They create an information extraction and retrieval system from unstructured natural language text documents to structured graphs along with natural language querying. They used the NLP techniques such as, semantic role extraction, phrase chunking, concept extraction etc to better understand input query and generate elements that constitute the graph. The Precision, Recall and Average accuracy of the system was 85.45%, 86.28% and 80.1% respectively.

Dwaipayan Roy et al[16] proposed a word embedding base query expansion technique for Information Retrieval on Wikipedia documents. They used two models, i) Word2Vec ii) fastest – used sub word information for learning. The similarity between the word is calculated with Jaccard similarity. The query terms are matched with embedded word vectors using Indexing Unit Composition (IUC) method. The MAP for Word2Vec and fast Text of the system is evaluated as 0.23 and 0.24 respectively. Also, they conclude that Word2Vec works well on stemmed collection and fast Text on unstemmed collection.

Shomi Khan et al [17] attempted for improving answer extraction for Ban- gali Question Answering system. In their work, demonstrated a wed document hierarchy and wordnet for answer retrieval using semantic matching with Anaphora-Catephora-resolution. Wordnet is referred to as a lexical database. The average accuracy of the system is observed as74%.

In a paper, Bo Xu et al [18] proposed a novel query expansion framework based on learning-to-rank methods for biomedical information retrieval. They used a term ranking model to select most relevant term for a query. In or uder to train the model they proposed a pseudo relevance feedback method. To refine the expansion terms, define and extract both the corpus-based term features and the resource-based term features to represent the terms as feature vectors, which are taken as the inputs for learning-to-rank methods to learn the term- ranking models. Different approaches to learning-to-rank are investigated for training the term-ranking models. The Average MAP of the system evaluated as 0.35. In another work [28], they proposed to optimize the pseudo-relevance feedback method, a classic query expansion method, using learning-to-rank methods to refine the set of expansion terms.

In a work Vaishali Singh et al [29] proposed a personalized approach to question answering using end-user modeling. According to the user information and interest area. Personalization of retrieved data can be performed using different similarity measures, such as attribute values similarity, entity values similarity, etc. The average precision and recall of the system are evaluated as 0.7 and 0.6 respectively.

Manasamithra P et al [20] proposed a method for convert natural language query to system understandable query using a hybrid approach. Which include keyword-based and semantic-based methods by using an efficient data structure-B-tree to store keywords which act as a knowledge base. The semantic analysis is carried out by using dependency parser. The system has experimented with an employee database. The analysis has shown that the execution time reduced by almost 86% while using B-tree.

In a work, Weiguo Zheng et al [21] proposed a semantic question answering system over knowledge graphs. They use a novel systematic method to under- stand natural language questions using a large number of templates by exploiting the knowledge graph and a large text corpus. The templates are executed by using semantic graphs. To select the target templates, use Semantic Dependency Graph (SDG). Perform entity level and structural level disambiguation during the conversion of natural language queries to structured queries. Finally, a SPARQL query can be constructed, then the corresponding answer will be returned. They conduct the study with Wikipedia text corpus- Dbpedia and freebase. The average precision of Dbpedia and freebase are 84.67% and 82.19%respectively.

Sheetal S et al [22] presented a novel method for calculating the similarity of documents using a graph model. They performed a modified method over WordNet and Wikipedia. In this approach, the weighted conceptual graph of the coexistence term is used for representing text documents. They used the co-reference resolution method to find the association of feature terms and weighted terms for graph construction. The semantic similarity is calculated by Wu-Palmers [31] method. Then an inverted index is created. The graph similarity is calculated by using vertex cosine similarity. The experiment is conducted on a news group dataset. The precision and recall ranges from 0.8to 1and F-measure is evaluated as 0.64.

Swathilakshmi Venkatachalam et al [23] proposed an ontology-based information extraction and summarization system for Tamil news content retrieval for users queries. By using Information Extraction, retrieve certain information from natural language and submit it to ontology. The ontology is clustered into two different domains. Then a multi-document text summarizer creates an overview of important events. Finally, the query extractor extracts data from the database and submits it to users according to their queries. The precision, recall, and F-measure of the system are evaluated as up to 90.1%, 88%, and 94% respectively.

Fan fang, Bo-wen Zhang, and Xu-cheng yin [24] developed a Semantic Se- quential Dependence Model (SSDM) for Biomedical article search, which is a combination of semantic information and the conventional Sequential Dependence Model (SDM). The synonyms are obtained automatically through the word-embeddings, here used word 2vec and skip-gram models. They used the neural network-based, SSDM language model. The create a thesaurus by using KNN classification algorithm. Afterwards, the query keywords are extracted and replaced with the synonyms from the thesaurus. Then the synonyms are used to generate possible sequences with the same semantics as the original query and these sequences are input into SDM to obtain the retrieved results.

Liang Pang et al [25] proposed a new deep learning architecture named Deep- Rank for relevance ranking in Information Retrieval. In their approach, they simulate human judgment process in relevance ranking. The relevance label is generated by three steps 1) relevant locations are detected by using a query- centric context 2) local relevance i.e. relevance between query and each query centric context is determined by using Convolutional-Neural Network (CNN) and two-dimensional gated recurrent units (2D-GRU)3) finally local relevance are aggregated by Recurrent Neural Network(RNN) to output a global relevance score. The Deep Rank model is trained by using the Stochastic Gradient Decent (SGD) method. The experiment is evaluated with LETOR 4.0 and large scale Chinese clicks trough data and the MAP for the same is evaluated as 0.49 and 0.41.respectively.

In a work, Ming Zhu et al [26] discussed the development of a neural network model for ranking documents for question answering in health care domain. The proposed model perform deep attention at word, sentence and document level. They also construct a large health care question-answering data set. They use a neural network model, HAR-a Hierarchical Attention Retrieval model for retrieving answers for health-related queries. The different components of the HAR model are 1) Word embedding-create a k-dimension word vector. 2) Encoder-use a bi-directional RNN (Bi-RNN) to encode the inter-document temporal dependencies within query and document words. 3) Compute the relevance of each query word w.r.t each word in the document by using a bi- directional attention mechanism.4) Query inner attention mechanism used to encode variable-length queries into fixed-size embedding by the self-attention mechanism. 5) Finally use a document hierarchical inner attention mechanism to get a fixed dimensional representation document by using sentence level embedding. Then they use a negative sampling mechanism for optimization of the results. They use health care data set and named it as Heaith QA. The MRR of the system is evaluated as 87.87% and recall as 96.84%.

Z huyin Dai et al[27] proposed a contextual neural language model-BERT, to provide deeper text understanding for Information Retrieval. BERT (Bi- directional Encoder Representation from Transformers) used for ad-hoc document retrieval. The input for BERT is the concatenation of query and documents tokens. Tokens are embedded then separate the query from document embeddings and added to token embedding. The position embedding is also added for word orders. The tokens are gone through several layers of trans- formations. At each layer, a new contextualized embedding is generated for each token by finding the weighted-sum of all other token embeddings by us- ing several-attention matrices. Words with strong attention are considered as more close to the target word. Then the output embedding of the first to- ken is used for all query-document pairs. It then inputs into a Multi-Layer Perceptron (MLP) to predict the relevance possibility. This can be augmented with search knowledge. They used two standard datasets-Robust-04-news corpus and Clueweb09-B. About the accuracy, the NDCG of Robust-04 and Clueweb 09-Bare 0.52 and 0.29 respectively. It is shown that BERT performs well on Robust-04 than Clueweb09-B dataset.

Y uan Zhang et al [28] developed a Graph Embedding-based ranking model for Product Search (REPS) for e-commerce search. The system integrated the click-graph features into a unified neural ranking framework. In their model, they first introduce a simple neural network architecture as the base model, then plugged a graph embedding technique for better retrieval performance. First, they represent terms of queries and product description as vectors.

Then input these vectors to CNN layers for semantic feature extraction, max- pooling layers are used for dimension reduction. Finally use Multi-Layer Perceptron (MLP) to transform semantic feature vectors into the same vector space as query and output relevance score. They used graph embedding during training phase using CNN or RNN. Evaluate the model using the CIKM Cup-2016 Track-2 data set. The MRR, MAP, NDCG of the model is evaluated as 0.49, 0.46, 0.53 respectively.

Navjot Kaur et al[29] developed a semantic information retrieval system in the music domain. They used string ontology for semantic Information Retrieval in which they performed reformulation techniques in order to implement the multilingual concepts. The maximum precision and recall of the system are evaluated as 83% and 72% respectively.

In a work, Ping Wang et al [30] proposed a deep learning based translate-edit Model for Question to SQL generation for Question Answering on Electronic Medical Records, which adapts the widely used sequence to sequence model to generate SQL query for a given query, and performs the required edits using an attentive copying mechanism and task specific look-uptables. They created a large- scale Question-SQL paird at a set, named MIMIC SQL from the publicly available Electronic Medical Records, it contains two sets, the first set contain template questions and the second consists of natural language questions. Finally Conducted an extensive set of experiments on MIMICSQL dataset for both template questions and natural language questions to demonstrate the effectiveness of the proposed model. They adopt an RNN sequence to sequence framework for the Question to SQL generation, the encoder reads a sequence of word embeddings of input tokens and turns the min to a sequence of encoder hidden states and the testing is performed with implement a beam search algorithm for the SQL generation. Their model gains a significant performance improvement on both development and testing dataset and 30 per cent, on average more accurate than other models. The average accuracy of the system was evaluated as0.97.

4. Comparative Study and Analysis of Different Models

The comparative analysis of various models discussed in the previous section is tabulated in Table1.

Table 1 contains the name of authors and paper, domain and language, methodologies used, and accuracy of reviewed research works. The different methodologies used are NLP techniques, similarity measures, Word Sence Disambiguation, word/ document embedding, graph embedding, and neural net- work methods such as CNN, RNN, etc. Also, the accuracy of the different models is included. The accuracy of the systems is evaluated by using different parameters like precision, recall, MAP, MRR, NDCA, and F-measures, etc.

Figure 2 shows the number of works used against different technologies, almost all works are used NLP techniques. Most recent works are based on Word embedding and Neural Network methods.

Table 1: Comparative study and Analysis of Different Models

Sl No	Authors	Domain	Language	Methods	Accuracy
1	Nadia Soudani et al[5]	Semantic space	Arabic	Word Sence Disambiguation	MAP-0.97 to 7.5
2	Shengxian Wan et al [6]	Yahoo Answers	English	LSTM,Bi-LSTM,MLP.	MV-LSTM- 11.4% more
3	Saravanakumar et al [7]	Open Domain	English	Latent Semantic Analysis	Precision- 0.77
					MRR - 0.79
4	Piyush Arora et al [8]	WebAP	English	Pseudo Relevance Feedback	NDCG-0.16 MRR-0.36
				Word Embedding	
5	Nouha Othman et al [9]	Yahoo-Webscope	English	Word Embedding	MAP- 0.39 - 0.45
				Cosine Similarity	
6	Shenghui Wang et al [10]	Medline	English	Document Embedding	Average recall
				Ariadine	Doc2Vec-93.3%,Ariadine-86.3%
7	Prathyusha et al[11]	Semantic Web	English	SPARQL,Ontology	-
8	Reshma PK et al[12]	University Data,	English	Ontology,DLQuery	Precision-87% , Recall-56%
9	Pratibha Bajpai et al[13]	Google, Bing	Hindi	Word Sence Disambiguation	MAP-
				Component Analyser	Google-0.45,Bing-0.35.
10	D Thenmozhi et al[14]	Agriculture	Tamil	Ontology	MAP-95.36%
				Word Sence Disambiguation	
11	Piyush Mital et al [15]	Wikipedia	English	Wikipedia	Precision-85.45%
				Semantic Role extraction	Recall-86.28%
12	Dwaipayan Roy et al[16]	Wikipedia	English	Word Embedding-	MAP-Word2Vec-0.23,
				Word2Vec,Fast Text	Fast Text-0.24
13	Shomi Khan et al[17]	Bangali database	Bangali	Anaphora Catephoraresolution	Avg.Accuracy-75%
14	Bo Xu et al [18]	TREC genomics	English	Pseudo-relevance feedback	Avg.MAP-0.35
				Learning-to-rank	
15	Manasamithra P et al[19]	Employee data	English	Dependency parser	Time reduced
				B-tree	Time reduced-86%
16	Weiguo Zheng et al[20]	Wikipedia	English	Semantic Dependency	Average precision-
		Dbpedia,Freebase		Graph,SPARQL	Dbpedia-84.67,Freebase-82.19
17	Swathilakshmi et al[23]	News dataset	Tamil	Ontology	Precision-90%,Recall-88%
					F-measure-94%
18	Fan fang et al[24]	MEDLINE	English	Word embedding-	MAP-0.34
				Word2Vec,skip-gram	
19	Liang Pang et al[25]	LETOR4.0,	English	CNN,RNN,	MAP-LETOR4.0- 0.49
		Chinese Click		2D-GRU	Chinese Click-0.41.
20	Ming Zhu et al[26]	HealthQA	Englush	Word embedding,	MRR-87.87%,
				MLP	Racall-96.84%
21	Zhuyun Dai et al[27]	Robust-04	English	Word embedding,	NDCG-Robust-04- 0.52
		Clueweb09-B		MLP.	Clueweb09-B-0.29
22	Yuan Zhang et al[28]	E-commerce data-	English	Graph Embedding	MRR- 0.49, MAP- 0.46
		CIKM Cup-2016.		CNN,RNN,MLP.	NDCG- 0.53
23	Navjoth et al[29]	Music dataset	Multi-ligual	String-Ontology	Precisio-83%,Recall-72%
24	Ping Wang et al[30]	EMR	English	TRanslate-Edit Model	Avg. accuraccy-0.9
				LSTM,RNN.	

From the literature, it is clear that most of the semantic Information Retrieval works are done in the field of English, a few works are done in native languages such as Arabic, Tamil, Bengali, etc. NLP techniques, Machine Learning, and neural network techniques are used for Natural Language Processing. Ontology and word/ document embedding used for document modeling and Machine Learning and Neural Network methods are used for document retrieval.

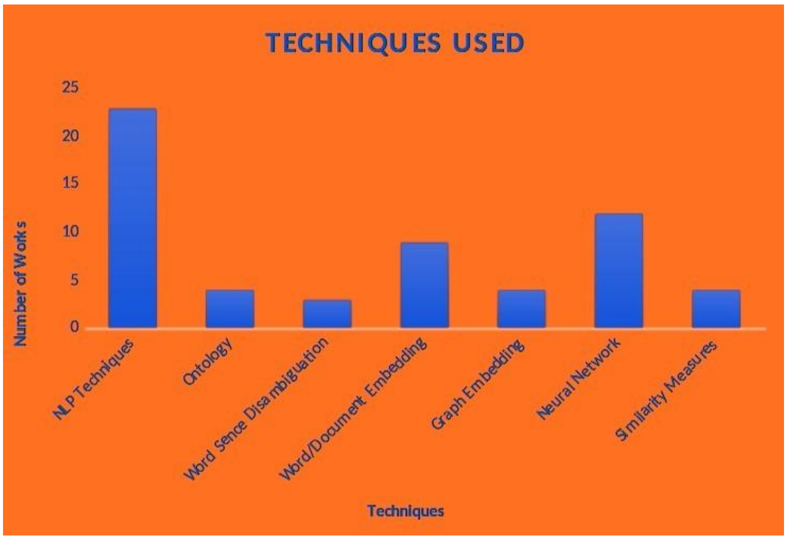

Figure 2: The number of papers used with different technologies

5. Conclusion

The semantic level Information Retrieval systems are used for retrieval of relevant answers for natural language queries from unstructured natural language datasets .The different methodologies used are NLP techniques, Machine Learning, and Neural Networks methods. Most recent works are based on context level embedding and Neural network methods. Also, most of the semantic Information system exists now are in English languages, a few research works are done in native languages like Tamil, Kannada, and Arabic, etc. There are few Information Retrieval works are done in Malayalam language [2][3], as now, almost all of the mare keyword based. No effective semantic Information Retrieval system exists in Malayalam language.

6. Direction for Future Work

Now we are going to propose a semantic level Malayalam Question Answering system for answers health-related queries. Malayalam is an agglutinative and morphologically rich language. Due to the complexity, the development of an Information Retrieval System for Malayalam is a tedious and time-consuming task. Although the system becomes very helpful for people, especially the un- educated, who seek answers to their health-related queries.

References

[1] Green BF, Wolf AK, Chomsky C, and Laughery K. ”Baseball: An automatic question answering”. In Proceedings of Western Computing Conference, Vol.19, 1961, pp. 219–224.
[2] Proceedings of AFIPS Conference, Vol.42, 1973, pp. 441–450
[3] Arjun Babu, Sindhu L. “An Information Retrieval System for Malayalam Using Query Expansion Technique”, IEEE, 978-1-4799-8792-4/15, 2015. Archana S.M., Naima Vahab, Rekha Thankappa, C. Raseek, ”A Rule-Based Question Answering System in Malayalam corpususing Vibhakthi and POSTag Analysis”, International Conference on Emerging Trends in Engineering, Science and Technology, ICE TEST 2015.
[4] Liji S K and Lajish V L, “An Efficient Malayalam Query Processing System for University Enquiry”, Proceedings of the Eighth National Conference on Indian Language Computing (NCILC 2018), CUSAT, Kerala, March2018
[5] Nadia Soudani, Ibrahim Bounhas, Yahya Slimani.”Semantic Information Retrieval: A comparative experimental study of NLP Tools and Language Resources for Arabic”, IEEE, DOI 10.1109/ICTAI.2016.133,2375-0197/16,2016.
[6] Shengxian Wan, Yanyan Lanall,” A Deep Architecture for Semantic Matching with Multiple Positional Sentence Representations”, Proceedings of the Thirtieth AAAI Conference on ArtificialIntelligence@2016.
[7] Saravanakumar Kandasamy and Aswani Kumar Cherukuri, “information Retrieval for Question Answering System Using Knowledge-Based Query Reconstruction by Adapted Lesk And Latent Semantic Analysis”, Research Gate, International Journal of Computer Science and Applications·January2017.
[8] PiyushArora, Jennifer Foster, and Gareth J. F. Jones.”Query Expansion for Sentence Retrieval Using Pseudo Relevance Feedback and Word Embedding”. Springer International Publishing. DOI: 10.1007/978-3-319-65813-1 8, 2017.
[9] Nouha Othman, Rim Faiz and Kamel Smaili.”A Word Embedding based Method for Question Retrieval in Community Question Answering”. ICNLSSP-International Conference on Natural Language, Signal and Speech Processing, ISGA, Dec2017.
[10] Shenghui Wang and Rob Koopman. ”Semantic embedding for information re- trieval”Workshop on Bibliometric enhanced Information Retrieval.,BIR2017
[11] Prathyusha Kanakam, S. Mahaboob Hussain and D. Suryanarayana.”HPL Algorithm For Semantic Information Retrieval with RDF And SPARQL”. Research Article.Jr. of In- dustrial Pollution Control33,2017.
[12] Reshma PK, Lajish VL. ”Ontology-Based Semantic Information Retrieval Model for University The domain” International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 15 2018 pp. 12142-12145, Number 152018.
[13] Pratibha Bajpai, Parul Verma and

SyedQ.Abbas."English-Hindi Cross-Language Information Retrieval System: Query Perspective". Pratibha Bajpaietal./Journal of Com- puter Science. DOI:10.3844/jcssp.2018.705.713,2018.

[14] D Thenmozhi and Chandrabose Aravindan. "Ontology-based Tamil–English cross-lingual information retrieval system". Indian Academy of Sciences. S˚adhan˚a, https://doi.org/10.1007/s12046-018-0942-7,2018

[15] Piyush Mital, Saurabh Agrawal al."Graph-based Question Answering System", IEEE, 978-1-5386-5314-2/18/©2018.

[16] Dwaipayan Roy, Debasis Gangulyal."Using Word Embeddings for Information Retrieval: How Collection and Term Normalization Choices Affect Performance". 3269206.3269277 CIKM '18, ACM.,Torino, Italy, October2018

[17] Shomi Khan, Khadiza Tul Kubra, Md Mahadi Hasan Nahid, "Improving Answer Ex- traction For Bangali Q/A System Using Anaphora-Cataphora Resolution". International Conference on Innovation in Engineering and Technology (ICIET), IEEE, 27-29 December, 2018, 978-1-5386-5229-9/18/0©2018.

[18] Bo Xu, Hongfei Lin, Yuan Lin. "Learning to Refine Expansion Terms for Bio-medical Information Retrieval Using Semantic Resources".10.1109/TCBB..2801303,IEEE/ACM, 2018

[19] Aishali Singh, Sanjay K. Dwivedi. "Personalized approach for automated ques-tion answering in restricted domain". International Journal of Information Technology, Springer. https://doi.org/10.1007/s41870-018-0200-6.2018.

[20] Manasamithra P, H. C. Vijayalakshmi "NLP for Information Retrieval using B Trees". International Journal of Computer Applications (0975 – 8887).Volume 182 – No.5, July 2018

[21] Weiguo Zheng, Jeffrey Xu Yu, Lei Zou, Hong Cheng. "Question Answer- ing Over Knowledge Graphs: Question Understanding Via Template De-composition". Proceedings of the VLDB Endowment, Vol. 11, No. 11.Au- gust. 2018 DOI: https://doi.org/10.14778/3236187.3236192.2018

[22] Sheetal S. Sonawane , Parag Kulkarni. "Concept based document similar- ity using graph model" International Journal of Information Technology, Springer. https://doi.org/10.1007/s41870-019-00314-w.2019.

[23] Swathilakshmi Venkatachalam , Lakshmana Pandian Subbiah et al[3]."An ontology- based information extraction and summarization of multiple news articles". International Journal of Information Technology, Springer.2019.

[24] Fanfang, Bo-wenzhang, and Xu-chengyin."Semantic Sequential Query Expansion for Bio-medical Article Search", 2169-3536 IEEE, 2018.

[25] Liang Pang, Yanyan Lan et al." Deep Rank: A New Deep Architecture for Relevance Ranking in Information Retrieval". arXiv:1710.05649v2[cs.IR],ACM,22Jul2019.

[26] MingZhu,AmanAhujaetal."AHierarchicalAttentionRetrievalModelforHealthcare Question Answering"., IW3C2 (International World Wide Web Conference Committee), ACM ,ISBN 978-1-4503-6674-8/19/05.https://doi.org/10.1145/3308558.3313699,2019

[27] Zhuyun Dai, Jamie Callan. "Deeper Text Understanding for IR with Contextual Neural Language Modeling". Association for Computing Machinery. ACM ISBN978-1-4503-6172- 9/19/07, https://doi.org/10.1145/3331184.3331303,©2019.

[28] Yuan Zhang, Dong Wang, Yan Zhang. "Neural IR Meets Graph Embedding: A Ranking Model for Product Search". The Web Conference, San Francisco, CA, USA. ACM ISBN 123-4567-24-567, May2019.

[29] Navjot Kaur, Himanshu Aggarwal[4]."Query reformulation approach using domain specific ontology for semantic information retrieval". International Journal of Information Technology,Springer.https://doi.org/10.1007/s41870-020-00464-2.2020.

[30] Ping Wang, Tian Shi, Chandan K. Reddy. "Text-to-SQL Generation for Question An- swering on Electronic Medical Records".IW3C2 (International World Wide Web Confer- ence Committee).,ACM ISBN978-1-4503-7023-3/20/04,2020.

[31] Wu Zetal."Verbssemantics and lexical selection".In:Proceedingsofthe32ndannual meeting on Association for Computational Linguistics. Association for Computational Linguistics.

[32] Keet Sugathadasa, Buddhi Ayeshaal."Legal Document Retrieval using Document Vector Embeddings and Deep Learning". Computing Conference 2018 10-12 July2018.

[33] Xu, B., Lin, H.,Lin,Y.(2016)."Assessment of learning to rank methods for query expansion". Journal of the Association for Information Science and Technology, 67(6): 1345-1357.2016.

[34] Tom Young, Devamanyu Hazarika, Soujanya Poria, Erik Cambria. "Recent Trends in Deep Learning Based Natural Language Processing", arXiv, IEEE Computational intelli- gence magazine,Nov2018.

[35] T. Kawamura, K. Kozaki, T. Kushida, K. Watanabe, and K. Matsumura, "Expanding science and technology thesauri from bibliographic datasets using word embedding," in Proc. IEEE Int. Conf. Tools Artif. Intell., Nov.2017.

[36] Shickel, B., Tighe, P. J., Bihorac, A, Rashidi "Deep EHR: A Survey of Recent Ad- vances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis", JBHI.2767063, IEEE, 2017.

[37] Jimmi, Guido Zuccon and Bevan Koopman."Knowledge Graphs and Semantics in Text Analysis and Retrieval". Information Retrieval Journal. https://doi.org/10.1007/s10791- 018-9344-z. Springer Nature B.V.2018.

[38] Sumit Kumar Mishra, V.K. Singh. "Building Semantic Information Retrieval System For Legal Cases From Heterogeneous Adapted And Diverse Data Sources Using Extended GAIAMethodologyForMulti-AgentSystem",byIEEE.978-1-5090-6785-5/18/©2018.