

Automated Generation of Question Answering System using Semantic Web

Suguna. M, Priyanga B, Prakash D.

Abstract: Question answering system provides answer to user's question according to the requirement. Question answering used in Information Retrieval processing, Natural Language, Artificial Intelligence (AI), Document Retrieval, Automatic evaluations. In QA system, once the question is posted by the client, the system need to find meaning of the words such as synonyms and provide correct answer to the user. It is very difficult to find answer from large collection of information available in Internet. The process involved in QA is question generation, answer filtering and store in the database. The proposed approach is used to retrieve the answers for the posted query in an efficient manner and reduce time consumption. WAD approach is used to analyze the answer and rank the accuracy of the answer with the existing method.

Keywords: Answer Ranking, Natural Language Processing, Cloud, Named Entity Recognition, Information Retrieval.

I. INTRODUCTION

Internet plays a very important role for current technological changes which paved the way for people to access and share the information across the global world. World Wide Web has the large database with million web pages. People make use of Search engine companies like Google and Bing in order to receive better results. Many users found very difficult to read their content through links/URL so they may be looking for straight forward answer for their posted query.

To solve this problem, question answering system has been developed. User requires a meaningful search which is known as semantic web. QA system which requires Named Entity Recognition, Disambiguation of entities, and analysis of the semantic question, expanding user query and query execution. Question Analysis will be done in client side and answer is retrieved from Amazon Web Service Cloud Database. QA System can be implemented using natural language, ontology based QA and semantic web based QA.

A. Natural Language

Natural language in artificial intelligence is concerned by making computers and human languages to get interacted to process huge amount of data. NLP is used for paragraph extraction, Webpage's, community mining, extracting hyponyms and key word. The goal of Natural language in QA is used to receive query from the user and provide an adequate answer for the posted query.

Manuscript published on 30 November 2018.

*Correspondence Author(s)

Suguna.M, Assistant Professor-II, CSE Department, Kumaraguru College of Technology, Coimbatore (Tamil Nadu), India.

Priyanga.B, PG Student, CSE Department, Kumaraguru College of Technology, Coimbatore, (Tamil Nadu), India.

Prakash D, Research Scholar, Anna University, Chennai (Tamil Nadu), India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

B. Ontology-based QA

Ontology a vocabulary, which make the queries to get stretched and submit them to Google search engine. WordNet ontology checks for vocabulary in sentence such as noun, verbs, synonyms, antonyms, adverbs and articles.

C. Web-based QA

QA System found answers from web resources for retrieving answers for posted questions by user. Web resources consist of large amount of information in order to provide accurate answer but it consumes more time. The objective of QA system is to analyze the question then search and finally generate the answer.

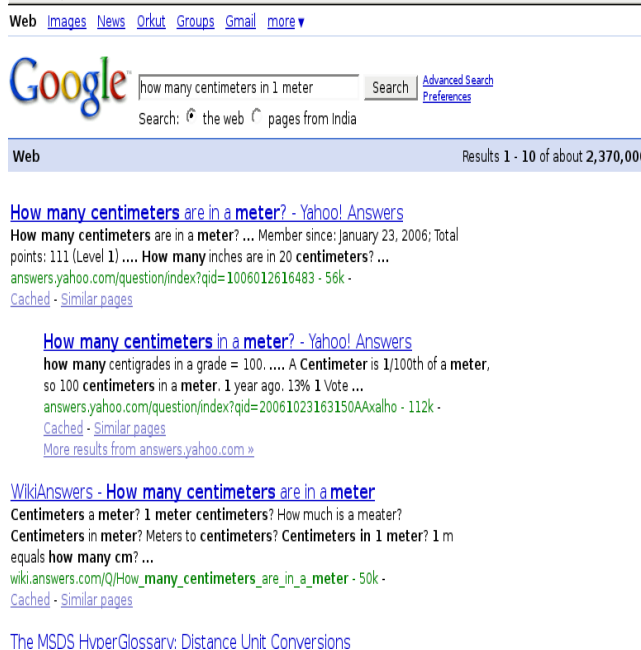


Fig.1. Web-Based QA

II. METHODOLOGIES

The process used in this QA system is:

A. Query Generation

Receiving the query from the user and identifying the type of question to provide accurate answer.

B. Query Pre-processing

Question Pre-processing which searches for the answer from documents and predicts correct answer and provides result.

C. Expansion of user Query

Query expansion was developed to identify similarity measure in posted query by using named-entity recognition.



D. Word Net

Wordnet is an English database which consists of adjectives, adverbs, nouns, phrase, verbs, synonyms and antonyms. This kind of database helps in matching of words in semantic web.

E. Ontology

The ontological words like an, are, and, how, when, which are removed to get semantic keyword. All these words are stored in an knowledge database.

F. Web search

It fetches accurate answer from the relevant document for the posted query.

G. Filtering of Answers

WAD algorithm uses Natural language for ranking the document with the existing document to find similarity between the words.

H. linking of Entities

Entity Link consists of variety of snippets, code and queries that need to be identified for particular users.

I. Ranking of Answers

Once the answer is retrieved it is ranked with the help of semantic similarity between words.

J. Precise Answer

The answer need to be validated based on probability conditional model and Web-semantic driven Algorithm.

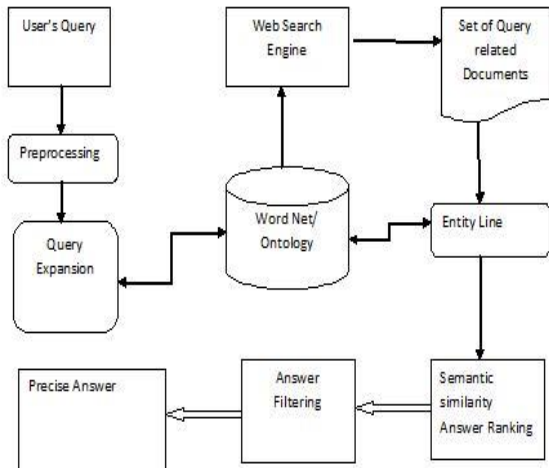


Fig. 2. Flow of WAD Methodology

III. WAD ALGORITHM

For each Qi

Step1. Query Preprocessing → Convert user query into NLP Question format for current input.

Step2.Expansion of user Query → Snippets Selection {identify the ontology} Semantic Relation {Synonyms, Antonyms, Hyponyms}

Step3. Linking of Entities → Remove the irrelevant documents.

Step4.Precise answer →To avoid ambiguity information retrieval.

Step5. Ranking → Head Word, NLP, Final word

Step6. Ranking of Answers → Calculate link distance and threshold limit for final word

For each scored candidate sentence do
Step7. Answer Type Prediction → use head word and query words

Step8. Validating user answer → Identify Question Type (Qt) and Information Selection (Ai).

IV. ANALYSIS OF QA SYSTEM

A. Accuracy Value

The accurate value depends on number of correct answers to the related tested value from the dataset.

$$\text{Accuracy} = \frac{\text{Total number of right answers}}{\text{Number of Related tested values}}$$

B. Precision Value

The value of Precision depends on the total no of right questions and the no of testing samples.

$$\text{Value of Precision} = \frac{\text{Number of right Questions}}{\text{Number of testing samples}}$$

C. Recall Value

The Recall Value is the no of questions retrieved and the no of answers.

$$\text{Recall} = \frac{\text{Number of Questions Retrieved}}{\text{Number of Answers}}$$

D. F- measure Value

F-measure Value is the ratio of precision value and recall combined to retrieve the answer.

$$\text{F-measure value} = 2 * \frac{\text{Precision Value} * \text{Recall Value}}{\text{Precision Value} + \text{Recall Value}}$$

V. ANALYSIS OF RESULT

It shows that accuracy is increased for WAD when compared with PSQALD for various combinations of datasets and bar chart is produced.

Table 1: Performance Analysis Table

Total Number of Questions	Algorithm	Percentage	Precision value
100 – 500	PSQALD	78%	Lower
	WAD	88%	Higher
500 -1000	PSQALD	55%	Lower
	WAD	65%	Higher
1000 – 1500	PSQALD	78%	Lower
	WAD	88%	Higher
and above	WAD	88%	Higher

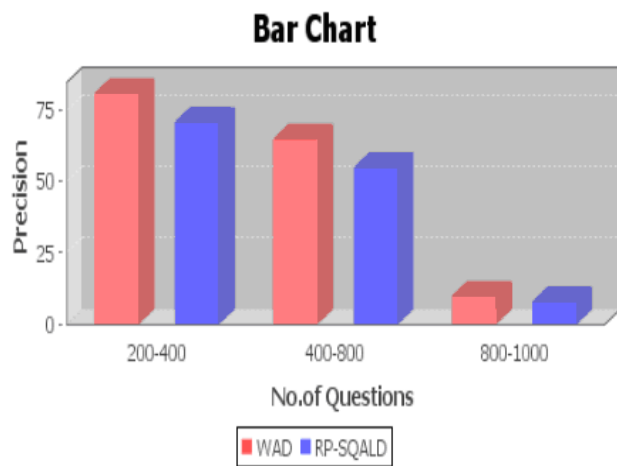


Fig. 3. Analysis Result Based on Performance

In the above bar chart the result is analyzed by calculating both WAD and PSQALD algorithm on Question Answer Dataset by changing the huge number of input questions and answers. The precision value gets reduced in PSQALD when compared with WAD based on testing samples.

VI. CONCLUSION

The objective of this work is to provide correct answers for the posted query. The proposed system will validates the user answer and provides the precise answer. PSQALD algorithm is measured based on probability, performance analyses gets reduced. Compared with other QA system, WAD approach produce more accurate and attains high ranking.

REFERENCES

- West, R., Gabrilovich, E., Murphy K, Sun, S., Gupta, R.: Lin D, Knowledge base completion via search-based question answering. In: ACM Proceedings of the 23rd international conference on World wide web, pp.515-526, (2016).
- Rini Wongso, Meiliana, Derwin Suhartono.: Question Answering System using Named Entity Recognition. In: 3rd International conference on Information Tech., Computer, and Electrical Engineering (ICITACEE), pp.19-21, (2016).
- MdMoinul Hoque, Paulo Quaresma: A Content-Aware Hybrid Architecture for Answering Questions from Open-domain Texts. In: 19th International Conference on Computer and Information Technology, pp.18-20, (2016).
- Pooja, P., Walke, Shivkumar Karale: Implementation Approaches for Various Categories of Question Answering System. In: IEEE Conference on Information and Communication Technologies, pp.8-15, (2015).
- Soricut, R., Brill, E.: Automatic question answering using the web: Beyond the factoid. In: Journal of Information Retrieval-Special Issue on Web Information Retrieval, Vol.9, pp.191-206, (2014).
- Petr Sosnin: Question-Answer Modeling in Conceptual Design of Automated Systems. In: IEEE MELECON, pp.16-19, (2012).
- Lopez, V., Uren, V., Motta, E., Pasin, M., AquaLog: An ontology-driven question answering system for organizational semantic intranets. In: Web Semantics: Science, Services and Agents on the World Wide Web, Vol.5, No.2, pp.72-105, (2013).
- Lally, A., Prager, M., McCord, M., Boguraev, B., Patwardhan, S., Fan, S., Fodor, R., Chu-Carroll, J.: Question analysis: How watson reads a clue. In: IBM Journal of Research and Development, Vol.56, No.3.4, pp.2-1, (2012).
- Monahan, S., Lehmann, J., Nyberg, J., Plymale, J., Jung, A.: Cross-lingual cross-document conference with entity linking. In: TAC 2011 Workshop, (2011).
- Hirschman, L.; Gaizauskas, R.: Natural Language question Answering. In: ACM proceedings of the Natural Language Engineering vol7, No.4, pp 275-300, (2010).

- Suchanek, F., Kasneci, G., Weikum, G.: Yago: a core of semantic knowledge. In: ACM proceedings on WWW, pp.697-706, (2011).
- Babanne, V., Patil, S., Joshi, D.: Intelligent Question answering System. In: International Journal of Scientific & Engineering Research, Volume 4, Issue 5, (2013).
- Web site: www.searchdocs.net.
- Web site: www.cs.cmu.edu/QA-data.