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 straightforward 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.

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.

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.
Automated Generation of Question Answering System using Semantic Web

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.

For each scored candidate sentence do
Step7. Answer Type Prediction  \(\rightarrow\) use head word and query words
Step8. Validating user answer  \(\rightarrow\) 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.

\[
\text{Total number of right answers} = \frac{\text{Number of Related tested values}}{\text{Number of testing samples}}
\]

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.

\[
\text{F-measure value} = 2 \times \frac{\text{Precision Value} \times \text{Recall Value}}{\text{Precision Value} + \text{Recall Value}}
\]

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

<table>
<thead>
<tr>
<th>Total Number of Questions</th>
<th>Algorithm</th>
<th>Percentage</th>
<th>Precision value</th>
</tr>
</thead>
<tbody>
<tr>
<td>100 – 500</td>
<td>PSQALD</td>
<td>78%</td>
<td>Lower</td>
</tr>
<tr>
<td></td>
<td>WAD</td>
<td>88%</td>
<td>Higher</td>
</tr>
<tr>
<td>500 -1000</td>
<td>PSQALD</td>
<td>55%</td>
<td>Lower</td>
</tr>
<tr>
<td></td>
<td>WAD</td>
<td>65%</td>
<td>Higher</td>
</tr>
<tr>
<td>1000 – 1500 and above</td>
<td>PSQALD</td>
<td>78%</td>
<td>Lower</td>
</tr>
<tr>
<td></td>
<td>WAD</td>
<td>88%</td>
<td>Higher</td>
</tr>
</tbody>
</table>
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