International Journal of Engineering & Technology, 7 (3.6) (2018) 379-381

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Exploring the Web and Semantic Knowledge-Driven Automatic Question Answering System

S. Jayalakshmi[1*], Ananthi Sheshaayee[2]

[1]Research Scholar, Periyar University.
Asst. Professor, VISTAS.
[2]Associate Professor, Quaid-e-Millath Government College for Women (Autonomous), Chennai. E-mail:Ananthi.research@gmail.com
**Corresponding author E-mail:Jayalakshmi.research@gmail.com*

Abstract

The growth of information retrieval from the web sources are increased day by day, proving an effective and efficient way to the user for retrieving relevant documents from the web is an art. Asking the right question and retrieving a right answer to the posted query is a service which provide by the Natural Language Processing. Question Answering System is one of the best ways to identify the candidate answer with high accuracy. The web and Semantic Knowledge Driven Question Answering System (QAS) used to determine the candidate answer for the posted query in the NLP tools. This method includes Query expansion techniques and entity linking method to identify the information source snippets with ontology structure, also ranking the sentences by applying conditional probability between query and Answer to identify the optimal answer from the web corpus. The result provides an exact answer with high accuracy than the baseline method.

Keywords: *Semantic, syntactic, question answering, ontology, entity linking, conditional probability.*

1. Introduction

The searching of information on the web is increased day by day as the growth of web searching is rapid the demand of information retrieval with efficient and effective and easiest way is required [1]. The question answering (QA) system provides the best solution for relevant information retrieval instead of bulk of relevant documents [2].

Features in Question Classification

Lexical Analysis: process of converting the sequence of characters into meaningful information it includes stemming words, bigrams, and headwords of the questions.
Semantic analysis: provides the meaning of the posted words, it includes hypernyms, POS and WordNet are used to assign semantic meaning of the given words.
Syntactic analysis: Includes grammatical information about the sentence, Head word, POS, nouns, Verbs are essential for Question Classification.
Natural Language Processing (NLP): NLP is a field of artificial intelligence it deals with automatic understanding of natural languages; NLP plays a vital role in converting the human language into the machine understandable formal representations [3].
Question Answering system have inherited many techniques from machine learning and Natural Language Processing to retrieve precise answers automatically. The questions are classified into fine grained classes using a hierarchical classifier. The classifier gives the layered semantically related answers. [4]. The Lexical, Semantic and Syntactic features are used to improve the accuracy of Question classification. The question type contain abbreviations, description, entity linking, human idea, location, numeric and set if fine grained classifiers to evaluate the question and retrieve the relevant Answer for the given query.

2. Web and Ontology Based QA System

Web Based QAS: The QA(Question Answering) system provides the answer from the web corpus [5]. It is a dynamic information source to provide the accurate answer to given query. [6]. It predict the appropriate answer type of the question by using Machine Learning algorithm to validate the answer types and also assign the rank to answer by applying probabilistic graphical model[7].
Ontology Based QAS: it used to expand the query before submitting into the web search engine. The ONKI selector function used to expand the Query with a search interface [8]. The template based QA system generate the SPARQL[9] template method to identify the structure of the Question. The structured questions are submitted to the NLP engine for matching the Question and related Answer. To linking the entities of search result with the ontology to evaluate the accuracy of the candidate answer.

3. Overview of Web and Semantic–Driven Automatic QAS (WSDA)

The WSDA approach aims to provide the appropriate, concise answer automatically to Natural Language questions from the user posted through NLP Engine by applying entity linking method, used to search information from web corpus and Yago ontology.

The Semantic QA system extract information from ontology to reduce disambiguate of the questions. It concentrates on entity linking with multiple knowledge based features. To mapping the user question to ontology for retrieving the correct pattern of the text.

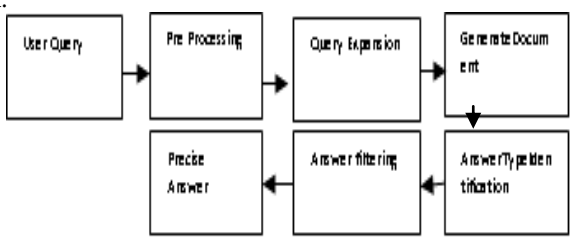

Fig. 1: WSDA methodology

Methodologies

User Query: Receive the User Query from the web through Natural Language Engine for further processing.
Pre Processing: To convert unstructured question pattern into structured pattern for question Answer process. It involves tokenization, snippets selection, stop word selection, stemming process are used to expand the Query
Query Expansion: used to identify the query term and novel term from a snippet. The snippet expansion is entered into Wordnet Ontology to obtain the sematic terms. The semantic terms contains synonyms, hypernyms and hyponyms of original text.
Generate Documents: The expanded query terms are used to give input to the web search engine to retrieve the set of relevant documents. From the set of documents the WSDA approach selects only the snippets of relevant documents which make an entity linking with ontology to retain the relevant candidate answer
Answer Type Identification: The WSDA approach identifying and examining the answer type. To validate the candidate answers for identify the answer accuracy. The WSDA approach focuses on user query with the WH-operator and linked all the information as the answer type appearing in a Question type. The informal Questions are determined by applying ontology.
Answer Filtering: Filer the answer which is relevant to the posted query, this process is used to eliminate the irrelevent and unwanted documents from the web, used to predict the appropriate answer for the question.
Precise Answer: To determine the answer type for each posted query by estimating the probability of set of query terms with all suitable answer types and validate the answers according to the document weightage and sentence formation.

WSDA System Process

User question

"What is the Largest Country in the world?"

Generating candidate answer

Preprocessing

Input: Natural Language Processing Question:
Output keyword: {What, Largest, Country, World}

Query Expansion

Key word identification, use search engine to retrieve top results in terms of snippets.

Search results

S1: Australia is the largest island and the smallest continent in the world
S2: Canada is the second largest country in the world
S3: Russia is the largest country in the world
The WSDA selects the snippets based on the threshold value that is the distance between the query term and novel term which less than the number of keywords in the query. The query contains 3 words, hence the novel term based on the threshold limits 3. From the search results s1and s2 are removed and s3 is determined as relevant answer for the posted query it will appear at the top of the list in the search result on the web.

4. Result and Discussion

The results are compared with the baseline approach RP-SQALD [10]. The dataset [11] [12] contains the QA pairs from the year 2008 – 2010, it contain the manually generated factoid questions and answers from ontology source.

Accuracy with Response Time and Question Length

The Comparative result of the WSDA and RP-SQALD approach are discussed here. The accuracy is determined with the length of question and response time. The RP-SQALD approach prolongs the response time by 81.6% when the question length from 2 to 10, but the WSDA approach consumes the response time by 20.93%. This method only focus on the snippets of documents and applies the probability based entity linking method while generating the expected answer. This approach balance the response time until reaches the certain limit on the question length and then slightly delays the response time.

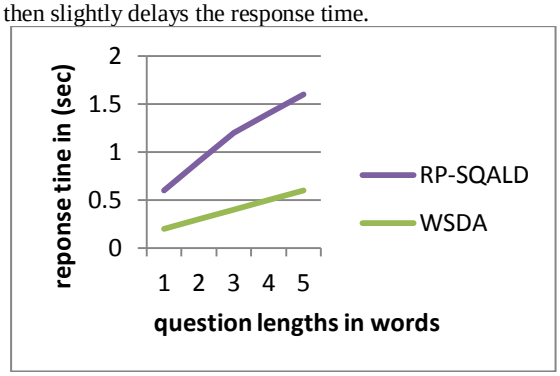

Fig. 2: Response time Vs question words

5. Conclusion and Future Direction

WSDA Approach managing the complexity in the QA system with the help of web and semantic relevance. The candidate answer sentences are ranked using the conditional probability based semantic score between the query and answer selection. The proposed approach accurately validates the recognized answer type with the list of relevant documents and identify the significant information as a precise answer, compared the baseline QA system, the WSDA approach attained the high accuracy.
To refine the algorithm in future to provide the better and accurate result for the large and descriptive question pattern, it gives high accuracy as a result for factoid question not for the descriptive, hence to refine and give high accuracy for both large and descriptive type question as the future work.

References

[1] Etzioni O, "Search needs a shake-up", *Nature*, Vol.476, No.7358, (2011), pp.25-26.
[2] Kolomiyets O & Moens MF, "A survey on question answering technology from an information retrieval perspective", *Elsevier*

transaction on Information Sciences, Vol.181, No.24, (2011), pp.5412-5434.

[3] Daud SP & Ribeiro CHC, "NLP–LEXICAL ANALYSIS APPLIED TO REQUIREMENTS", *Proceedings of the 9th Brazilian Conference on Dynamics Control and their Applications Serra Negra, SP-ISSN*, (2010), pp.2178-3667.

[4] Loni B, "A Survey of State-of-the-Art Methods on Question Classification", *Literature Survey Published on TU Delft Repository*, (2011).

[5] Brill E, Dumais S & Banko M, "An analysis of the AskMSR question-answering system", *EMNLP*, (2002), pp.257-264.

[6] West R, Gabrilovich E, Murphy K, Sun S, Gupta R & Lin D, "Knowledge base completion via search-based question answering", *ACM Proceedings of the 23rd international conference on World wide web*, (2014), pp.515-526.

[7] Ko J, Nyberg E & Si L, "A probabilistic graphical model for joint answer ranking in question answering", *ACM Proceedings of the 30th annual International SIGIR conference on Research and development in information retrieval*, (2007), pp.343-350.

[8] Tuominen J, Kauppinen T, Viljanen K & Hyvönen E, "Ontology-based query expansion widget for information retrieval", *Proceedings of the 5th Workshop on Scripting and Development for the Semantic Web, 6th European Semantic Web Conference*, Vol.449, (2009).

[9] Yahya M, Berberich K, Elbassuoni S, Ramanath M, Tresp V & Weikum G, "Natural language questions for the web of data", *EMNLP-CoNLL*, (2012), pp.379-390.

[10] Hakimov S, Tunc H, Akimaliev M & Dogdu E, "Semantic Question Answering System over Linked Data using Relational Patterns", *ACM Proceedings of the Joint EDBT/ICDT Workshops*, (2013), pp.83-88.

[11] http://searchdocs.net/

[12] http://www.cs.cmu.edu/~ark/QA-data/