Survey on Question and Answering Retrieval System using Hadoop

Atul Kumar Mishra[1], Prof. Amit Saxena[2]

[1]Department of Computer Science & Engineering,
Truba Institute of Engineering & Information Technology, Bhopal Madhya Pradesh, India
[2]Professor, Department of Computer Science & Engineering,
Truba Institute of Engineering & Information Technology, Bhopal Madhya Pradesh, India

---***---

Abstract - *Question and Answering System is main area of research in Natural Language. Key tasks of Question and Answer system gives exact answer of question which give by user. Question and Answering scheme is categorized into three group are open domain, closed domain and restricted domain. Using progressive Natural Language Processing device we will be developed a framework for inquiry replying scheme. In this paper we work on limited area question answering scheme. Proposed system work on keyword and question matching and return precise answer of question.*

Key Words: Natural Language processing, information retrieval, semantic similarity, limited domain, answer abstraction, reply ranking.

1. INTRODUCTION

While the set of papers which are recovered by the search engine cover a lot of info near the search topic but it may or may not contain exactly that information which the user is looking for [1].The simple knowledge after the question answering scheme is that the users just have to enter the question and the system will retrieve the most suitable and detailed answer for that question and arrival it to the user. Hence in those cases where the user is observing for a short and precise answer, question answering System plays a countless role quite than Search Engines, which usually provide big set links of those web sheets which might comprise the answer of that question. A characteristic Question Answering group can be separated into 3 modules namely: Question Dispensation module, Document Processing or Material Retrieval module and Answer Processing module.

Contains several sub modules and these modules use several Natural Language Dispensation Methods in instruction to extract the proper answer. The traditional Inquiry Replying scheme is planned to reply modest queries like "Who", "what", "when", "where", etc. But the present QA study attentions on reaching the scheme to answer composite questions, rapid questions, opinion questions etc. The paper guides a Inquiry Replying scheme that answers simple factoid, wh-questions by using a practice called Semantic Role Category.
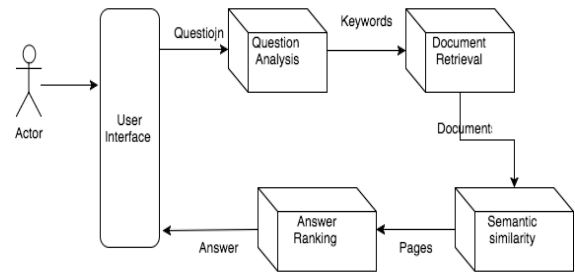

Figure 1. Block Diagram Question Answering System

The respite of the paper is prearranged as trails. The following segment defines the general building of a Query Replying Society. Section 3 debates particular of the related altogether in this part. The future construction is considered in section 4. The paper positions with the new setup and consequences.

1.1 ARCHITECTURE OF A QUESTION ANSWERING

In this unit we label the building of our organization. The general construction of the arrangement container be split into three main modules: (1) pre-processing, (2) question template matching, and (3) answering. Each module is described in point in next subsections.

Question Answering Schemes can be classified on foundation of the areas over which it has been constructed.

- Open Area Question Answering
- Close Field Question Answering
- Limited Province Question Answering

Open area question answering systems are domain independent. It relies on general ontology and world information. Frequently these schemes have a large group of data from where the compulsory answer is to be originating out. Meanwhile in situation of Open Domain question answering info satisfied is not of particular domain it can answer questions of numerous fields though here deep reasoning is not possible [3].

Close area inquiry replying schemes contract by queries in a detailed domain [3]. LUNAR and BASEBALL are the example of close domain QA systems .In this case the data set contains a very partial quantity of focused and organized material . Hereafter in situation of close area query answering

arrangements deep reasoning is probable but the problematic with these administrations was that due to the very minor size of data set they are not more than a 'Toy Systems"[4].

Education in restricted-domain speeches difficulties connected to the grouping of part- exact genetic factor into current state-of-the-art QA technology with the hope of achieving bottomless reasoning capabilities and reliable precision recital in real ecosphere requests. In fact, as a not too-long-term vision.

2. LITERATURE SURVEY

In most of the research papers [4, 5, 6] LUNAR [7] and BASEBALL [8] have been discussed as the earlier developed question answering systems. However there are numerous review replying administrations which have been developed with dissimilar perceptions later the knowledge of QA Structure has been coined

In a system advanced Attire P. M, Et.al [10], presented an architecture of ontology-based domain-specific accepted philological enquiry rejoining that affronts semantics and domain knowledge to improve both inquiry building and answer extraction.

Another system developed by Pragisha K. Et.al [11], labeled about the. It receives Malayalam natural verbal queries after the user and extracts most appropriate answer by analyzing a collection of Malayalam documents. The system handles four each question.

Research and appraisals in inquiry replying scheme industrialized by Sanjay K Dwivedi Et.al[12] propose taxonomy for symbolizing Question Answer (QA) systems, survey of major QA systems described in poetry and provide a qualitative analysis of them.

Table [I] offerings judgment about diverse kinds of system related to it [22].

S. No	Category of Question and Answering System	Question and Answering System Methods
1	Multilingual Question/Answering	Tokenization and pos tagging., Word sense disambiguation, Answer type identification, Keywords expansion, Semantic Disambiguation
2	Analysis of the Asks Question-Answering System	Query Reformulation, NGram Mining, N-Gram Filtering, N-Gram Tiling.
3	Multilingnality, Spatial- temporal context awareness, Textual entailment	Answering architecture
4	A Query Replying Scheme founded on Substantial Recovery and Validation	Expected Answer Type, Named Entities Presence,
5	A Cross Query Replying Scheme founded on Info Recovery and Response Authentication	Module, Hypothesis Generation Module, Document Processing and Indexing
6	A specifiable part bilingual Question	Answering architecture

Query replying is a problematic procedure of info recovery branded by material wants that are at smallest rather spoken as natural linguistic Pattern Corresponding Involuntary Replying Scheme For natural languages questions planned by Pachpind Priyanka Et.al [17], Frequently Asked QA Scheme that answers by pre- stowed responses to operator queries requested in even English, comparatively than keyword or ruling construction founded retrieval devices.

3. PROPOSED SYSTEM

Meanwhile together the Exposed Area QA Scheme and Close Area QA Scheme have their own pros and cons a original idea of Query Replying has been conceived by Molla& Vice do [4] called RESTRICTED DOMAIN QA SYSTEM, which is the central of these two parts.

We are influenced that investigation in limited domains will drive the meeting among structured knowledge-based and free text-based question answering.
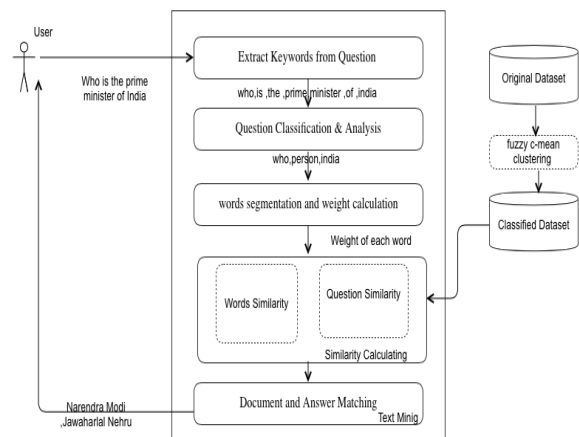

Figure 2. Proposed System

Our proposed system performs following operations:

1. Question Processing: Given Question is processed to get some important information from it. Steps through which question Processing Module passes and their descriptions are given below. Steps in Question Processing Module:
   a. Find the Type of given question using WH word.
   b. Find out the expected type of answer.
   c. Get the Keywords from the Question.
   d. Find out the Focus of the question.

The QA System is the Question Processing or Question Classification module. Various information, which we will get through this module, are the Type of Question, Expected Answer Type, Focus or Head Word of the Question and the Question Keywords.

WH word	Question Type		
	Factoid Type	**Definition Type**	**Descriptive Type**
Question	Who When What Where Which	How What Why	What

2. Document Dispensation: When query has continued treated we will change near the text dispensation unit. In this component the papers which are pertinent to the given query are saved and treated. Following steps used in document processing.

   a. Get the question in hand and search relevant documents using a reliable search engine.
   b. Take top relevant documents.
   c. Extract the content from these documents.
   d. Save these contents in to file

3. Answer processing**:** This module presents algorithms for extracting the potential answer for all the three categories of questions that is Definition Type of Question, Descriptive Type of Question and Factoid Type of Question.

4. Dataset Clustering: cluster dataset using fuzzy c-mean algorithm then process for question and answer processing.

4. CONCLUSIONS

In this paper we have proposed a outline for limited area question Answering System by progressive NLP tools and software. This basis can be castoff to change a Question Answering Scheme for extracting exact and exact response

from restricted domain written data set. The proposed framework not only delivers a simple and implementable outline for developing question Answering Classification but also provides a proper flow of data for answer extraction.

Since the future perfects everything over keywords and headword and is sovereign of the question or sentence structure, it has summary the overhead of question normalization. Moreover since the framework is given for restricted domain, it also handles the issue of word sense disambiguation. The major problem which exists with the proposed framework is that its performance is dependent on the act of hunt locomotive and the used NLP tools.

REFERENCES

[1]. Sreelakshmi V, Sangeetha Jamal, Survey Paper: Question Answering Systems, in National Conference on Computing and Communication - (NCCC), March 2014, GEC Idukki.

[2]. M Ramprasath, S Hariharan Improved Question Answering System by semantic reformulation, IEEE- Fourth International Conference on Advanced Computing, 2012.

[3]. Ali Mohamed Nabil Allam, and Mohamed Hassan Haggag, the Question Answering Systems: A Survey, International Journal of Research and Reviews in Information Sciences (IJRRIS), September 2012 Science Academy Publisher, United Kingdom

[4] Molla D., and Vicedo J., "Question answering in restricted domains: An overview", Computer Linguist, ppAI-6 1, 2007

[5] Moreda P., Llorens H., Saquete E., & Palomar M., "Combining semantic information in question answering systems", Information Processing & Management, pp.870-885, 20 1 1.

[6] Svetlana Stoyanchev, and Young Chol Song, and William Lahti, "Exact Phrases in Information Retrieval for Question Answering", Coling 2008: Proceedings of the 2nd workshop on Information Retrieval for Question Answering (IR4QA), pp. 9- 16 Manchester,UK. August 2008".

[7] Woods W.A, Kaplan R.A, Nash-Webber.B, "The lunar sciences naturallanguage information system" , Final report: BBN Report #2378. Technical report, Bolt Beranek and Newman Inc., Cambridge, MA., June 1972.

[8] Green RF, Wolf A.K., Chomsky, K. Laughery, "BASEBALL: An automatic question answerer", in: Proceedings of Western Computing Conference, vol.19, pp. 2 19-224, 196 1.

[9] Ittycheriah A, Franz M, Zhu WJ, Ratnaparkhi A and Mammone RJ. IBM"s statistical question answering system. In Proceedings of the Text Retrieval Conference TREC-9, 2000.

[10] Athira P. M., Sreeja M. and P. C. Reghuraj ⬚Department of Computer Science and Engineering, Government Engineering College, Sreekrishnapuram, Palakkad, Kerala, India, 678633. Architecture of an Ontology-Based Domain-Specific Natural Language Question Answering System. ⬚

[11] Pragisha K. "design and implementation of a QA system in Malayalam". ⬚

[12] Sanjay K Dwivedi, Vaishali Singh. Research and reviews in question answering system Department of Computer Science, B. B. A. University (A Central University) Luck now, Uttar Pradesh, 226025, India. ⬚

[13] Poonam Gupta, Vishal Gupta ⬚Assistant Professor, Computer Science & Engineering Department University Institute of Engineering & Technology Panjab University, Chandigarh. ⬚

[14] Kolomiyets, Oleksander. And Moens, Marie-Francine. "A survey on question answering technology from an information retrieval perspective". Journal of Information Sciences 181, 2011.5412-5434. DOI: 10.1016/j.ins.2011.07.047. Elsevier. ⬚

[15] Moreda, Paloma., Llorens Hector., Saquete, Estela. And Palomar, Manuel."Combining semantic information in question answering systems" Journal of Information Processing and Management 47, 2011. 870- 885. DOI: 10.1016/j.ipm.2010.03.008. Elsevier. ⬚

[16] Ko, Jeongwoo., Si, Luo., and Nyberg Eric. "Combining evidence with a probabilistic framework for answer ranking and answer merging in question answering" Journal: Information Processing and Management 46, ⬚2010 541-554. DOI: 10.1016/j.ipm.2009.11.004. Elsevier. ⬚

[17] Pachpind Priyanka P, BornareHarshita N, KshirsagarRutumbhara B, Malve Ashish D BE Comp S.N.D COE & RC, YEOLA," An Automatic Answering System Using Template Matching For Natural Language Questions".