

QUESTION ANSWERING SYSTEMS: ANALYSIS AND SURVEY

Eman Mohamed Nabil Alkholy, Mohamed Hassan Haggag and Amal Aboutabl

Faculty of Computers & Information, Helwan University, Cairo, Egypt.

ABSTRACT

computing environment In real world that using a computer to answer questions has been a human dream since the beginning of the digital era, Question-answering systems are referred to as intelligent systems ,that can be used to provide responses for the questions being asked by the user based on certain facts or rules stored in the knowledge base it can generate answers of questions asked in natural , so this survey paper provides an overview on what Question-Answering is and its system ,as well as the previous related research with respect to approaches that were followed.

KEYWORDS

Question-answering, Neutral language processing, Answer Extraction, Evaluation Metrics.

1. INTRODUCTION

As technology and human-computer interaction advances, there is an increased interest in affective computing and a large amount of data is generated and made it available every day, and for that it requires to integrate and query a huge amount of heterogeneous data, for that NLP has been recognized as a possible solution that capable to manipulate and represent the complex query as uncertain and complicated that are existing in them. Which leads to the generation of QA consisting of Equation and Answer that mapping between these information [1]. However, Question Answering (QA) is a fast-growing research area that combines the research from Information Retrieval, Information Extraction and NLP. It can be seen as the next step in information retrieval is to automatically generating answers to natural language questions from humans, that allow users to pose questions in natural language and receive succinct answers.

2. BACKGROUND

Over the past four decades Question Answering systems have been transitional much at par with the whole of natural language processing. In this section, we present a previous work on development of QA system and it's propose, the earliest system was developed in 1959 (in the spirit of the era called The Conversation Machine), and A large number of QA system have been developed since 1960's One of the most memorable systems was BASEBALL developed by (Green et al. 1961 in NL DB systems) [2]. Although, capable of answering rather complex questions, BASEBALL was, not surprisingly, for answering questions about baseball games played in the American league over one season, restricted to questions about baseball facts, and most question answering systems were for a long time restricted to front-ends to structured databases. And in 1963 they develop QA system PROTOSYNTHEX that permit user to ask a question in English, it accept statements in (sub set of English) as input to its database and accepts quotations as a query to the database. And for read and solve the kind of word problems

they develop (Problem-solving systems) STUDENT QAs system in 1964 That can read and solve the kind of word problems found in high school algebra books. The program is often cited as an early accomplishment of AI in natural language processing [3].

In early 1966 they provide QA system ELIZA that can communication with user) And This conversation can be simulated with a psychologist. It was able to converse on any topic by resorting to very simple rules that detected important words in the person's input, And in 1972 they develop SHRDLU that offered the possibility to operation of a robot in a toy world (the "blocks world)to be simulated with ability to ask the robot questions about the state of the world SCHOLAR QA system develop in 1973 it was a formal theory characterizing the variety of plausible inferences people use to ask questions about which their knowledge is incomplete. it has methods embed to lessons learned from such research into the SCHOLAR system [4].

In the same year 1973 they develop the first automatic question answering system (SAM). After three years in 1976 TRIPSYS (HWIM) was the first QA that understand speech Question , TRIPSYS(HWIM) was developed as the context for a research project in continuous speech understanding- it understands and answers questions about planned and taken trips, travel budgets and their status, costs of various modes of transportation to various places it's called HWIM (for "Hear What I Mean[2]). And the second famous QA system was (NL DB systems) Problem-solving systems LUNAR in 1977[5] for answering questions about the geological analysis of rocks returned by the Apollo moon missions.

In 1977 develop two QA system the first one is GUS it was a dialog system for airline reservation second one was LIFER that develop to asking questions about U.S. Navy ships [6] This system used a semantic grammar with domain information built within in 1978 they start to develop system that deal with story comprehension (NL DB systems) QUALM an application that use for story comprehension and this application responsible for scripts and plans in a very restrictive domain. In 1983 Salton and McGill describe question answering (QA) systems as mainly provide direct answers to questions. Finally Kupiec (1993) employed similar but rather simpler WH-question models to build a QA system [8].

In 1991 was QA system (LILOG) text understanding system that operated on the domain of tourism information in a German city.in 1993 they start to combined NLP with the use of an online encyclopedia by developing MURAX QA system that combined NLP with the use of an online encyclopedia with hand-coded annotations to sources. In subsequent developments, QAS aimed on making linguistic analysis of the questions to capture the intended requirements in a natural way [7]. In subsequent developments, one of QAS aimed to capture the intended requirements in a natural way IS to making linguistic analysis of the questions.

In recent 90's, question answering achieved a great progress due to the introduction of the Text Retrieval Conferences (TREC) question answering track there has been great progress in open domain Question answering (Voorhees 2001) . These systems use unrestricted text as a primary source of knowledge. One such system, MASQUE (ANDROUTSOPOULOS et al., 1993) use logic representation to represents natural language questions, and for retrieving intended information from database, it use to translates the logic query into a database query. It separates from mapping process the task of linguistic process. FAQ Finder (Burke et al., 1997) does matching of the question list that compiled in a knowledge base with questions through statistical similarity and semantic similarity and for syntax-based natural language understanding technique. In 1999 was LASSO that win the question answering task. It used question classification technique and syntax-based natural language understanding technique. Another QAS in (2000) was developed by Riloff

and Thelen (QUARC), that has the ability to classify questions into different types and use lexical and semantic clue to derive their expected answer [9] [10].

Later, the focus of developing QASs was shifted toward open domain QAS, TREC Evaluation campaign which is taking place every year since 1999 to manage and query large volume of data and represent most of research in open domain question answering from unstructured data sources, and that lead to question-answering evaluations as a recent success started as part of the Text Retrieval Conference (TREC). The best systems are now able to answer more than two thirds of factual questions in this evaluation, it describes the results and the associated evaluation Methodology develop by Ellen Voorhees, the second paper, by Buchholz and Daelemans, explores the requirements for answering complex questions that have compound answers or multiple correct answers [11].

The third paper, by Lin and Pantel, describes a new algorithm to capture paraphrases that allow a more accurate mapping from questions to potential answers. The fourth paper, by Light, Mann, Rilo and Breck, describes experiments that systematically factor and assess question answering into component sub-problems. The first TREC evaluation campaign provides a list of 200 questions and a document collection. The answers were known to be present in the collections. The maximum lengths of answers were allowed to be 50 or 250 characters[12]. Systems were asked to give 5 ranked lists of answers.

In the next campaign, TREC-9 held in 2000, the number of questions and size of document collections were increased. In TREC-10 in 2001, a new complexity with respect to answers. The lengths of answers were reduced to 50 words [11]. In TREC- 11, held in 2002, systems were expected to give exact short answers to the questions. In TREC from 2002 to 2007, the list of questions, definition questions, and factoid questions were included in the evaluation campaigns. In TREC 2005 and TREC 2006 there was temporal question in addition to 75 topics that contains different type of Questions. In TREC 2007, there were a collection of documents that include collection of blogs, and progress competitions that lead to increasing documents collection complexity, and questions complexity, and that also effect to increasing answer evaluation strategies complexity.

In 2000 they start to develop QA system for Unix operating system Unix Consultant For answered questions pertaining to the Unix operating system [10]. This system had ability to accommodate various types of users by phrasing the answer, and it has its domain that contain a comprehensive hand-crafted knowledge base. In 2001 there was two QA systems the first one was INSIGHT question answering system which uses some surface patterns, wins the question answering task in TREC-10, and the second one was SiteQ use the density-based extraction method to retrieve related passages first and then extract the exact answers in them, which can greatly improve the extraction speed.

in 2002 STARTQA system was the first web-based QA system for English and in the same year 2002 was Answer bus QA system(ODQA) they develop this system to accepts questions in several languages (extend the answer extraction process from the local data source to the World Wide Web, which allows them to deal with large count of questions. After one year they develop QA system (ARANEA) it was first fully downloadable open-source Web-based factoid question answering system, and in the same year 2003 create a QA system AQUA was a sophisticated automatic question answering system, which combines natural language understanding technique, ontological knowledge, logical reasoning abilities and advanced knowledge extraction techniques [12].

In 2008 they start to develop a new QA system with different approach and it was a List questions ask for different instances of a particular kind of information to be returned. In 2012 they develop QA Systems that oriented to work with opinions rather than facts can also be included in this category. In 2014 was QAKIS by Cabrio et al and FREITAS14 develop by Freitas and Curry and also INTUI3 by Dima, in 2015 was develop three QA system first was HAKIMOV15 by Usbeck et al [12]. and second was QASYO by Hakimov et al. last. However, QA systems have developed over the past few decades until they reached the structure that we have nowadays.

2.1 IMPLEMENTATION APPROACHES FOR QA SYSTEMS

The basic aim of QA System is to provide correct and short answer with high accuracy to the user and There are many Approaches used in Question answering system based on different purpose [13]. This section will present an implementation approaches for various categories of QA System such as:

2.1.1 FIRST APPROACH

That relied on artificial intelligence (AI) it called Linguistic approach and it has the ability to build QA logics by using methods that integrated natural language processing (NLP) techniques and knowledge base. It used Linguistic techniques for formulating user's question into a precise query that merely extracts the respective response from the structured database, it Used to understand natural language text, linguistic & common knowledge Linguistic techniques such as such as (tokenization, POS tagging and parsing [14].

2.1.2 SECOND APPROACH

(rule-based approach) that rely on the (rule-based mechanism), they are built to identify question classification features. Quarc developed by Rilloff et al., and Cqarc [15] developed by Hao et al. Used to identify the question class by using semantic clues in question and for looking lexical they used heuristic rules.

2.1.3 THE THIRD APPROACH

(Statistical approach) gives the better results than other approaches user can use this approaches successfully to applied the different stages of a QA system, it independent of structured query languages and can formulate queries in natural language form. IBM's statistical QA system was based on the statistical model. This system utilized maximum entropy model for question/ answer classification based on various N-gram or bag of words features. Moschitti had used Rocchio and SVM text classifiers for question and answer categorization. Availability of huge amount of data on internet increased the importance of statistical approaches [16].

2.1.4 The Fourth Approach

Is Pattern matching that have the ability to replace the sophisticated processing involved in other competing approaches by using the expressive power of text patterns. Some of patterns matching QA rely on templates for response generation while most of of the patterns matching QA systems [17].

2.1.5 THE FIFTH APPROACH

Is the Surface Pattern based approach. This approach rely on an extensive list of patterns to extracts answers from the surface structure of the retrieved documents. It's automatically learning based pattern or it is human crafted. [18]. However, to identify any question answer it depend on

the basis of similarity between their reflecting patterns having certain semantics as well as their Methods, definition and Characteristics, Limitation and Aims for each QA system that was developed before which will be helpful for new directions of research in this area (Table 1).

QA system	Method & definition	Characteristics and Implementation Issues	Limitation
BASEBALL [2]	Description: answers English questions about the scores teams, locations, and dates of baseball games. Method: Analyzed the question, using-linguistic knowledge-, into canonical form.	Characteristics: Input sentences have to be simple, and not contain-sentential- connectives, such as (and, or).- The data are stored in a data base in attribute-value format. - questions-transformed into the same format, but in an automatic way. Implementation: system was implemented in the Lincoln Laboratory. And it relied on Linguistic approach.	- its domain of {baseball only} (Close domain) -, where the attribute-value structures can be uniform, and the types of questions are limited.
STUDENT [3]	Description: Correctly solving the algebra problem Method: was taken-as-a demonstration-that the system understood the written statement of the problem.	characteristics: -Understanding-language-requires-world-knowledge. -read and solved high school algebra word problems. Implementation : It is written in Lisp	- systems were limited by the amount of knowledge-they contained. Closed Domain
ELIZA [4]	Description: system that simulated a Conversation with a psychologist. Method: the computer can read messages typed on the typewriter and respond by writing on the same instrument.	Characteristics: ability to converse on any topic by resorting to very simple rules that detected important words in the person's input. Implementation: it was first implemented in the SLIP language, that use as an extension to FORTRAN but with better functionality to process doubly linked lists, and Its present implementation is on the MAC time-sharing system at MIT. And relied on Linguistic approach.	- (closed domain) the knowledge stored in the structured database was only capable of answering questions asked within the restricted domain.
LUNAR [5]	Description: Designed to enable a lunar geologist to conveniently access the chemical analysis data on lunar rock that was accumulating as a result of the Apollo moon mission". Method: The entries in the analysis table-specify-the concentration of some constituent in some phase of some sample.	characteristics: -The system contains two data bases: a 13,000-entry table off chemical and age analyses of the Apollo 11 samples. - able to answer 90% of the in-domain questions posed by working geologists, without prior instructions as to phrasing. implementation system is implemented in LISP. Parse English question into a data base query, Syntactic analysis via augmented transition network parser and heuristics and relied on Linguistic approach. That can analyze automatically with a transition network parser and translated into a data base query language.	-Sophisticated, with the syntax and semantics of questions - having a particular database could not be easily modified to be used with different databases. -closed domain
GUS [6]	Description: Simulated a travel advisor. Method: it attempt to explore the integration of already existing programming technology for a performance demonstration.	characteristics: had access to a restricted database of information about airline flights. - is a steady genetic algorithm with subpopulation support. Implementation: system implemented in MChART .it is a framework for implementing parsers. And relied on Linguistic approach.	-closed domain -the knowledge stored in the structured database was only capable of answering questions asked within the restricted domain.
MURAX [7]	Description: searched the encyclopedia for noun Phrases identified from the question. Method: exploited the phrase relations contained in the question to match questions with answer hypotheses.	Characteristics: -used the technology of robust shallow parsing. -use the Internet as a corpus - answers hypothesize noun phrases Implementation: the sentences of matching text are selected to confirm phrase relations implied by the question, rather than being selected solely on the basis of word frequency.	-lack of basic-information extraction support -open domain

QUALM [8]	<p>Description: designed to test a story understanding system.</p> <p>Method: it requires question to be parsed and a conceptual graph to be built, and it include two stages: understanding the question and finding the answer.</p>	<p>Characteristics: it able to answer questions about ideas not specifically mentioned in the texts.</p> <ul style="list-style-type: none"> - represented as a conceptual graph and text comprehension <p>Implementation: - it implemented as a language-independent question answering module, and integrated into other natural language processing applications</p>	<p>-hard to classify and complex system</p> <ul style="list-style-type: none"> - (closed domain) highly restricted in terms of the domain, or world, knowledge required and the genre of text covered
SHRDLU [9]	<p>Description: it's a research system that help researchers understand the issues involved in modeling human dialogue.</p> <p>Method: the system answers questions, execute commands, and accepts information in normal English Dialog.</p>	<p>characteristics: contain a combination of syntax, semantics, and reasoning.</p> <ul style="list-style-type: none"> - it uses semantic information and context to understand discourse and to disambiguate sentences. -it could search back further through the interactions. <p>implementation : it's written in Micro Planner and Lisp (a general-purpose, multi-paradigm programming language.</p>	<ul style="list-style-type: none"> -constrained to a simple block world. -restricted-domain - has a core database handwritten by experts
QAKIS [10]	<p>Description allows end users to submit a query to an RDF triple store in English and obtain the answer in the same language.</p> <p>Method: it composed of two main modules the query generator takes the user question as input, generates the typed questions, and then generates the SPARQL queries from the retrieved patterns.</p>	<p>Characteristics: huge amounts of available semantic data</p> <ul style="list-style-type: none"> -implement a relation-based match for question interpretation, to convert the user question into a query language <p>Implementation: QAKIS addresses the task of QA over structured Knowledge Bases (KBs) where the relevant information is expressed in unstructured form to implement a relation-based match for question interpretation, to convert the user question into a query language.</p>	Open domain
IBM's statistical QA system [16]	<p>Description: it's IBM statically question answer for TRCE 9.Is an application of maxim entropy classification for QA predication and name entity marking</p> <p>Method: maxim entropy</p>	<p>characteristics: classification based on various N-gram or bag of words features.</p> <ul style="list-style-type: none"> - Trained on crops of DSTS--use REASON <p>Implementation: Maximum Entropy Model for this (IBM's QA) system used a two-pass approach based on Okapi formula and expansion of queries based on TREC-9 QA corpus. -Statistical approach based</p>	Open domain
A Web Based Automatic QAs [19]	<p>Description: presents an answer extraction method which based on the calculation of sentence similarity between question and answer. (Sentence Similarity Model)</p>	<p>characteristics: using Web data resource as database for question answering system</p> <ul style="list-style-type: none"> - use answer correctness as our evaluation me <p>Implementation: Statistical models applied for question classification</p>	<ul style="list-style-type: none"> -Open domain -difficulty of finding answers to questions little data available in the target language in which to search for answers.
NLJDB [20]	<p>Description: it's a database system Engaged the user in dialogues it allows the user to access information stored in a database.</p>	<p>characteristics: used with large databases.</p> <ul style="list-style-type: none"> -could be configured to interface to different underlying database. Implementation: it was implemented entirely in Prolog. It transformed English questions into Prolog expressions, which were evaluated against the Prolog database. The code of Chat-80 was circulated widely) 	Closed Domain the knowledge stored in the structured database was only capable of answering questions in restricted domain.
Statistical Approaches for Answer-Finding [21]	<p>Description: investigates whether a machine can automatically learn the task of finding, within a large collection of candidate responses</p>	<p>characteristics - Usenet FAQ documents and customer service call-center dialogues from a large retail company</p> <p>Implementation - statistical models applied for question classification. using , adaptive tf.idf algorithm and automatic query</p>	Open domain

	Method: use four statistical techniques for answer-finding	expansion with latent variable model.	
PHLIQAI [22]	Description: it Designed to answer short questions against a data base containing fictitious data about computer installations in Europe and companies using them. Method: - use of the lambda calculus	Characteristics: Have three division of the translation from natural language questions to data base 1-English-oriented Fomal Language (EFL). 2- World Model Language (WML). 3-Data Base Language (DBL). Implementation: in PHLIQAI system Questions are translated into a formal language used to access the data base.	the limitation to a narrow domain
(LADDER) [23]	Description: it's a database systems It was designed as a natural language interface to a database of information about US Navy ships. Method: used semantic grammars to parse questions to query distributed database. It can developed as a prototype system for understanding questions posed in English about a naval domain.	Characteristics: used with large databases . - grammar must be tailor-made for each given database. - only support simple one table queries or multiple table queries . Implementation: used a semantic grammar to parse questions and query a distributed database. it translated each English question into one or more relational database queries, prosecuted the queries on a remote computer. Based on LIFER parser, which interpreted sentences according to a 'semantic grammar.'	-a different grammar had to be developed whenever Ladder was configured for a new application. - a narrow domain
Ask-Jeeves [24]	Description : allowed end-users to teach the system new words and concepts at any point during the interaction it's database systems method: The user stated requests in English, and Ask suitable requests to the appropriate underlying systems.	Characteristics: has its own built-in database. - use the Internet as a corpus Implementation : it's pointing the questioner to Web links that might contain information relevant to the answer to the question.	All the applications connected to Ask were accessible to the end-user through natural language requests -open domain
Answer Bus [25]	Description: is an question answering system based on sentence level Web information retrieval method : Answer Bus extracts sentences that are determined to contain answers.	Characteristics: High performance & accuracy for response time.-use the Internet as a corpus. -It accepts questions in several languages. Implementation: is based on sentence level information retrieval. It accepts users' natural-language questions in different language and extracts possible answers from the Web.	-it only extracts the named entities that match question types -open-domain
AQUA [26]	Description: allow users to ask questions in everyday language and receive an answer quickly and with a context Method: Use a similarity algorithm to map between names of relations in the knowledge base and names of relations in the ontology	Characteristics: it use semantic annotations to perform inferences . - AQUA's inference engine operates within the framework of multi-sorted logic, Implementation: it uses NLP technology, Logic and a hand-crafted ontology. And using Dice coefficient and WordNet algorithm. This algorithm is used to ensure that the question does not fail .	Closed domain
START [27]	Description: it's a First Web-Based Question Answering System, Designed To Answer Questions That Are Posed To-It-In-Natural Language. method: It Parses Incoming Questions, Matches the Queries Created from The Parse Trees Against Its Knowledge Base And Presents-The Appropriate-Information Segments To The User.	Characteristics Ability to Answer millions of English-Questions-- use the Internet as a corpus - used technique called "natural language annotation" Implementation consists of two modules. The understanding module: analyzes English text and produces a knowledge base that encodes information found in the text. -the generating module: produces English sentences. Used in conjunction with the technique of natural language annotation. - Linguistic based QA systems	-ambiguous modification linguistically-uninformed QA - have difficulty handling-semantic symmetry - open domain

3. QA SYSTEM OVERVIEW

Question answering is a process that understanding user natural language query and has the ability to provide a correct answer and extract it from retrieving relevant documents, data, or knowledge base .A question answering (QA) Question-answering systems are referred to as intelligent systems ,that can be used to provide responses for the questions being asked by the

user based on certain facts or rules stored in the knowledge base it can generate answers of questions asked in natural [28].

3.1 QA SYSTEM PURPOSE AND TYPES

The purpose of a QA system is to provide correct answers to user questions in both structured and non-structured collection of data. there are three main types for the QA Depending on the target domain and the way questions are answered it will present in table

Table 4: Types of QA Systems

TYPES OF QA SYSTEMS		ADVANTAGE
CLOSED-DOMAIN	<p>Built for very specific domains and exploit expert knowledge.</p> <p>The first QA systems of this type were developed in the 1960s. Any data not present in the database is usually considered out-of-domain.</p> <p>Closed-domain question answering deals with questions under a specific domain, This type of QA is easier for the vocabulary is more predictable, and ontologies describing the domain are easier to construct</p> <p>-Two of the most cited closed-domain QA systems are BASEBALL [2] and LUNAR [5]</p>	<p>Deals with very specific data which usually does not contain ambiguous terms and as a result can be processed more easily.</p>
OPEN DOMAIN	<p>Can be asked about virtually any topic and can theoretically extract the answer from any textual collection.</p> <p>- It can deals with unrestricted topics. Hence, questions may concern any subject. The corpus may consist of unstructured or structured texts.</p> <p>- Two of good example open domain QA systems are AnswerBus [25] and Askjeeve [24]</p>	<p>An alternative to search engines, available on the Web. Instead of providing a list of relevant keywords user just asks a question.</p> <p>-Automatically build queries and retrieve relevant documents, extract answers from the retrieved text snippets and present them as a confidence-ranked list to the user.</p>

4. QA SYSTEM ARCHITECTURE & ALGORITHM

4.1 SYSTEM ARCHITECTURE

As shown in figure (1) a QA system contain three main part beside other supplementary components: question classification (QUESTION PROCESSING), information retrieval (DOCUMENT PROCESSING) and answer extraction (ANSWER PROCESSING). The user writes a question using the user query interface. Then this query is used to extract all the possible answers for the input question. The architecture of Question Answering system is as shown in Figure1

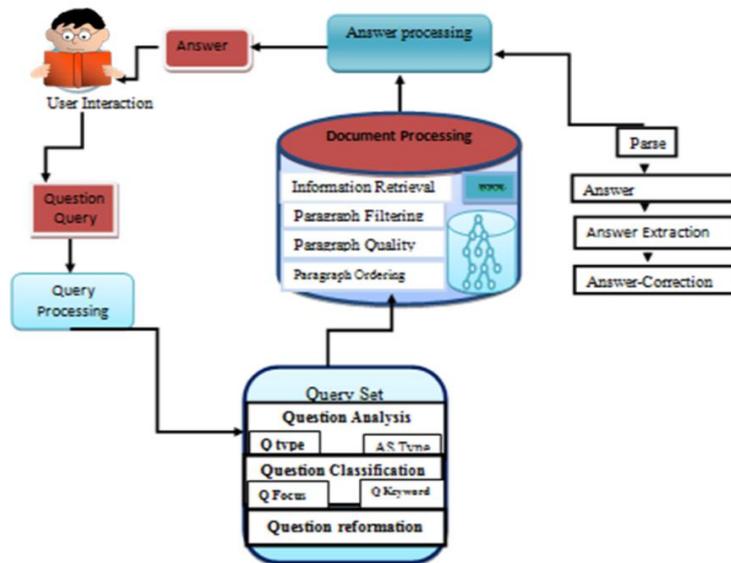


Figure (1): Question Answering System Architecture

4.1 QUESTION ANSWERING SYSTEM ALGORITHM

4.1.1 STEP ONE

(Question processing): Given a question as input, user writes his question by an interface the function of the question processing module is to process and analyses the question, and to create some representation of the information requested module is to process and analyses the question, and to create some representation of the information requested analyses the question, and to create some representation of the information requested. This leads to the classification of question that include three parts: Query Interface, question analyzer and Question classification. This step help to generate complex structured queries . and detects the expected answer type of a question e.g. the expected answer type of " When was last time that you feel this headache?" is date or "Do other family members have similar headaches?" Is yes/no this information helps guide the answer extraction process [29].

4.1.2 STEP TWO

After Input questions -Determining the question type: is selected from question taxonomy that system uses. and candidate answers are extracted using all the information gathered in the previous steps, e.g., keywords, entities and relations. The remaining candidate answers are then filtered and ranked according to different measures of similarity to filter out candidates with incompatible types. The Document Processing Module retrieves documents from the corpus that are likely to contain answers to the user's question. It consists of a query generation algorithm, text search engine and information retrieval that retrieve the relevant documents based upon important keywords appearing in the question. The query generation algorithm takes an input the user's question and creates a query containing terms likely to appear in documents containing an answer. This query is passed to the text in system, which uses it to retrieve answer. [30].

4.1.3 STEP THREE

The answer processing module is responsible for identifying, extracting and validating answers from the set of ordered paragraphs passed to it from the Document Processing Module, takes input from the retrieval component and tries to retrieve an exact phrase to return as an answer to achieve this required parsing and detailed question analysis by using of answer extraction algorithms. The identity answer extraction returns the CenterPoint of the passage, stripping words from either end until it fits within the specified answer. Then Answer Display The result and converted into required text which is required by the user and displayed to the user.

4.2 EVALUATION METRICS

There are several parameters that are used to analyze the performance of different Question Answering Systems. In this section we describe some of the evaluation metrics used in Question Answering System to evaluate its performance such as: precision, recall and F-measure, a weighted harmonic mean of precision and recall, can be defined with respect to predicates for the purposes of QA evaluation. These precisions and recall metrics express true precision and recall, not approximations, when coupled with an answer key in which the judgments can be reasonably assumed to be exhaustive [31].

4.2.1 PRECISION

Precision- Recall is the most common metric to evaluate information retrieval system. Precision is the ratio of retrieved documents that are relevant to all retrieved documents in the ranked list [30].

$$\text{precision} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{retrieved documents}\}|} \quad (1)$$

4.2.2 RECALL

$$\text{recall} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{relevant documents}\}|} \quad (2)$$

Recall in information retrieval is the fraction of the documents that are relevant to the query that are successfully retrieved

For example: text search on a set of documents recall is the number of correct results divided by the number of results that should have been returned In binary classification, recall is called sensitivity. So it can be looked at as the probability that a relevant document is retrieved by the query. It is trivial to achieve recall of 100% by returning all documents in response to any query [32]. Therefore, recall alone is not enough but one needs to measure the number of non-relevant documents.

4.2.3 F-MEASURE

A measure that combines precision and recall is the harmonic mean of precision and recall, the traditional F-measure or balanced F-score:

$$F = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (3)$$

known as the F measure, because recall and precision are evenly weighted. It is a special case of the general F_β measure (for non-negative real values of β):

$$F_\beta = (1 + \beta^2) \cdot \frac{\text{precision} \cdot \text{recall}}{\beta^2 \cdot \text{precision} + \text{recall}} \quad (4)$$

Two other commonly used F measures are the F_2 measure, which weights recall higher than precision, and the $F_{0.5}$ measure, which puts more emphasis on precision than recall.

The F -measure was derived by van Rijsbergen (1979) so that " F_β " measures the effectiveness of retrieval with respect to a user who attaches β times as much importance to recall as precision" [32].

$$E = 1 - \frac{1}{\frac{\alpha}{P} + \frac{1-\alpha}{R}} \quad (5)$$

It is based on van Rijsbergen's effectiveness measure

Their relationship is $F_\beta = 1 - E$ where

$$\alpha = \frac{1}{1 + \beta^2}. \quad (6)$$

6. CONCLUSIONS

In this survey paper are show an overview on what Question-Answering is and its architecture and how can be used as intermediate semantic representations of vague expressions. we pin down the previous related research with summarized and organized recent research results in a novel way that integrated and added understanding to work in the question-answering field. It emphasized the classification of the existing literature, developing a perspective on the area, and evaluating trends. However, because it is impossible for a survey to include all or even most of previous research, this survey included only the work of the top-publishing and top-cited authors in the QA field.

REFERENCE

- [1] Hendrix. G.G, Sacerdoti, E.D,sagalowicz. D. Slocum. J. "Developing a natural Language interface to complex data in ACM Transaction on database system. 3(2). pp. 105- 147,1978.
- [2] Green BF, Wolf AK, Chomsky C, and Laughery K. Baseball: An automatic question answerer. In Proceedings of Western omputing Conference, Vol. 19, 1961, pp. 219–224.
- [3] [Cook, 2006] Cook, J. L. (2006). College students and algebra story problems: Strategies for identifying relevant information. Reading Psychology, 27:95 – 125.
- [4] Weizenbaum J. ELIZA - a computer program for the study of natural language communication between man and machine. In Communications of the ACM, Vol. 9(1), 1966, pp. 36-45

- [5] Woods W. Progress in Natural Language Understanding - An Application to Lunar Geology. In Proceedings of AFIPS Conference, Vol. 42, 1973, pp. 441–450.
- [6] Bobrow, D. G., Kaplan, R. M., Kay, M., Norman, D. A., Thompson, H., and Winograd, T. (1977). GUS, A frame driven dialog system. *Artificial Intelligence*, 8, 155–173.
- [7] Julian Kupiec. 1993. MURAX: A robust linguistic approach for question answering using an on-line encyclopedia. In Robert Korfage, Edie Rasmussen, and Peter Willett, editors, Proceedings of the Sixteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 181–190. Special issue of the SIGIR FORUM
- [8] Lehnert, W. G. (1978). *The Process of Question Answering: A Computer Simulation of Cognition*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- [9] Huettner, A. (2000) “Question Answering”. In: 5th Search Engine Meeting.
- [10] Voorhees EM. The TREC-8 question answering track report. In Proceedings of TREC-8, 1999, pp. 77-82.
- [11] K.L.Kwok,L.Grunfeld,N.Dinstl,and M.Chan.TREC-9 Cross language,web and quationt -answering track experiments using PIRCS.In Voorhees and Harman .
- [12] E.M.Voorhees and D.K. Harman,editors,Proceedings of the Ninth Text REtrieval Conference (TREC-9)
- [13] Stefanie Tellex, "Pauchok: A Modular Framework for question Answering", Master Thesis Submitted to the Department of Electrical Engineering and computer science, Maccachusetts institute of Technology, June 2003..
- [14] Michael Kaisser. Question Answering by Searching Large Corpora with Linguistic Methods. Master's thesis, Saarland University, Germany, 2004.
- [15] Riloff E and Thelen M. A Rule-based Question Answering System for Reading Comprehension Tests. In ANLP /NAACL Workshop on Reading Comprehension Tests as.
- [16] 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.
- [17] Joho, H. (1999) "Automatic detection of descriptive phrases for Question Answering Systems: A simple pattern matching approach". MSc Dissertation. Department of Information Studies, University of Sheffield. Sheffield, UK
- [18] Deepak Ravichandran, Abraham Ittycheriah and salim roukos, "Automatic Derivation of surface text pattern for a maximum Entropy Based question answering system", Work done while the author was an intern at IBM TJ Watson research center during summer 2002.
- [19] Cai D, Dong Y, Lv D, Zhang G, Miao X. A Web-based Chinese question answering with answer validation. In Proceedings of IEEE International Conference on Natural Language Processing and Knowledge Engineering, pp. 499-502, 2005.
- [20] J.-L. Binot, L. Debillé, D. Sedlock, and B. Vandecapelle. Natural Language Interfaces: A New Philosophy. *SunExpert Magazine*, pages 67–73, January 1991

- [21] Berger A, Caruana R, Cohn D, Freitag D, and Mittal V. Bridging the lexical chasm: statistical approaches to answer-finding. In Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval, 2000, pp. 192-199.
- [22] Bronnenberg, W.J., Bunt, H.C., Landsbergen, S.P.J., Scha, RoJ.H., Schoenmakers, W.J., van Utter,n, E.P.C. (1979) The question answering system PHLIQAI. In L.Bolc (ed.), Natural communication with computers, McMillan, London; Hanser Verlag, M~nehen.
- [23] SACERDOTI, E.D. Language access to distributed data with error recovery. Proc. 5th Int. Joint Conf. on Artificial Intelligence, Cambridge, Mass., Aug. 1977.
- [24] Ask Jeeves. 1996. www.ask.com Site last visited in 28-march-2018.
- [25] AnswerBus, Question Answering System. Website: <http://answerbus.com> J. Allen. 1995. Natural Language Understanding. The Benjamin/Cummings Publishing Company, Menlo Park, CA.
- [26] Vargas-Vera M. and Motta E and Domingue J. (2003a): AQUA: An Ontology-Driven Question Answering System. AAAI Spring Symposium, New Directions in Question Answering, Stanford University, March 24-26, 2003.
- [27] Boris Katz, Sue Felshin, Deniz Yuret, Ali Ibrahim, Jimmy Lin, Gregory Marton, Alton Jerome McFarland, and Baris Temelkuran. 2015. Omnibase: Uniform access to heterogeneous data for question answering. In Proceedings of the 7th International Workshop on Applications of Natural Language to Information Systems (NLDB 2015).
- [28] Unmesh sasikumar, Sindhu L, "A survey of Natural Language question answering system", international journal of computer applications(0975-8887), volume 108 -No 15. December 2014 .
- [29] Anthony Fader, Luke Zettlemoyer, and Oren Etzioni. 2014. Open question answering over curated and extracted knowledge bases. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14, pages 1156–1165, New York, NY, USA. ACM
- [30] Sanjay K. Dwivedi and vaishali singh "Research and reviews in question answering system", International Conference on Computational Intelligence: Modeling Techniques and Applications CIMTA) 2013 Procedia Technology 10 (2013) 417 – 424.
- [31] Ellen M. Voorhees and Dawn M. Tice. 1999. The TREC-8 question answering track evaluation. TREC-8 Proceedings, pages 41–63, Nov.
- [32] VAN RIJSBERGEN, C.J. Information Retrieval, Second Edition, Butterworths, London (1979).